

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



## **TRABAJO FIN DE MÁSTER**

**Análisis transcriptómico de células del epitelio pigmentado de la retina con una mutación en PRPF8.**

**Máster Universitario en Bioinformática y biología computacional**

**Autor: Cabrera Alarcón, José Luis**

**Tutor: Dopazo González, Ana**

**Unidad de genética del Centro Nacional de Investigaciones Cardiovasculares**

**Ponente: Tejero Franco, Hector**

**Unidad de Bioinformática del Centro Nacional de Investigaciones Oncológicas**

**Madrid, Enero de 2018**



### *Agradecimientos*

*En primer lugar quiero agradecer a Ana Dopazo y Miguel Carballo por haberme dado la oportunidad de participar en este proyecto. También quiero dar las gracias a los compañeros de la unidad de bioinformática del CNIC, por la acogida recibida durante mi estancia allí y en especial a Manuel y a Fátima por haberme proporcionado una ayuda indispensable para dar forma al proyecto.*



<b>Índice</b>	
<b>1. Resumen</b>	<b>5</b>
<b>2. Objetivos</b>	<b>6</b>
<b>3. Introducción</b>	<b>7</b>
<b>4. Materiales y métodos</b>	<b>13</b>
4.1. Material biológico	13
4.2. Pipeline y software empleado	14
4.3. ASapp	17
<b>5. Resultados</b>	<b>20</b>
5.1. Análisis de expresión diferencial de genes	20
5.2. Análisis funcional	23
5.2.1. Análisis funcional en RPE	23
5.2.2. Análisis funcional en fibroblastos	28
5.3. Resultados <i>splicing</i> alternativo	28
5.3.1 Resultados de ASapp	28
5.3.1.1. Comparación paciente Vs control en células RPE	28
5.3.1.2. Comparación paciente Vs control en fibroblastos	30
5.3.1.3. Comparación fibroblasto Vs RPE en control	33
5.3.1.4. Comparación fibroblasto Vs RPE en paciente	35
5.3.2. Resultados de SwitchSeq	37
5.3.2.1. Comparación paciente Vs control en células RPE	37
5.3.2.2. Comparación paciente Vs control en fibroblastos	38
5.3.2.3. Comparación fibroblasto Vs RPE en control	39
5.3.2.4. Comparación fibroblasto Vs RPE en paciente	40
5.3.3. Resultados de DEXSeq	41
5.3.3.1. Comparación paciente Vs control en células RPE	41
5.3.3.2. Comparación paciente Vs control en fibroblastos	42
5.3.3.3. Comparación fibroblasto Vs RPE en control	43
5.3.3.4. Comparación fibroblasto Vs RPE en paciente	44
<b>6. Discusión</b>	<b>45</b>
6.1. Valoración de resultados del análisis de <i>splicing</i> alternativo	45
6.2. Valoración de los resultados del análisis de expresión diferencial y funcional	47
6.3. Valoración de la expresión diferencial de genes de la lista RP	48
6.3.1. Genes de la lista RP sobre-expresados	48
6.3.2. Genes de la lista RP reprimidos	50
6.4. Valoración comparada de ASapp, DEXSeq Y SwitchSeq	52
<b>7. Conclusiones</b>	<b>53</b>
<b>8. Bibliografía</b>	<b>54</b>

## **1. RESUMEN:**

La retinosis pigmentaria es un grupo de enfermedades genéticas que determinan la degeneración de la retina, causando ceguera. Muchas de las mutaciones que la producen están relacionadas con el splicing de pre-RNA. En este estudio analizamos datos RNAseq procedentes de un paciente con una mutación en el factor de splicing PRPF8. Se llevó a cabo un análisis *pairwise* de expresión diferencial, funcional para familias de genes y un análisis de *splicing* alternativo entre muestras de paciente y control, para muestras de células del epitelio pigmentario de retina y fibroblastos. Para la valoración del splicing alternativo entre condiciones se desarrolló el software ASapp, una herramienta escrita en R, con un interfaz gráfico generado con Shiny. ASapp se basa en el cálculo de la distancia de Jensen-Shannon para la detección de genes que experimentan splicing alternativo entre diferentes condiciones, en comparaciones pairwise. De acuerdo con lo observado en este estudio, la mutación en los pacientes determina una alteración de los patrones de splicing normales en células del epitelio pigmentario de la retina (RPE), reduciéndose la incidencia de retención intrónica. La alteración en los patrones normales de splicing genera un desajuste en la regulación de la expresión normal en estas células. La alteración en los perfiles de expresión en células RPE, conduce a una situación de estrés oxidativo, además de la disrupción de la función normal del retículo endoplasmático. Esto conlleva la activación del sistema de emergencia conocido como *unfolded protein response (UPR)* que a su vez implica la activación como mecanismo de defensa del sistema *Ubiquitin/Proteasome System (UPS)*. Adicionalmente, se activa la expresión de agentes que promueven la inflamación y la activación del sistema de complemento. Otra consecuencia de la alteración en la expresión normal, es que la ciliogénesis, el mantenimiento de estructuras ciliares y el transporte a través de ellas, se desajustan, siendo funciones capitales para la viabilidad de la retina. En síntesis, todos estos cambios concluyen en la disfunción de las células RPE, que son células auxiliares de

las que dependen los fotorreceptores, lo que provoca finalmente el mal funcionamiento y la muerte, primero de bastones y finalmente de conos.

## **2. OBJETIVOS**

1. Análisis de expresión diferencial de genes entre células de control y células de paciente.
2. Análisis funcional de genes diferencialmente expresados.
3. Desarrollo de una aplicación (ASapp) para la detección e interpretación biológica de eventos de *switching*.
4. Estudio de *splicing* alternativo mediante tres herramientas distintas, DEXSeq, SwitchSeq y ASapp.

### **3. INTRODUCCIÓN**

El término retinosis pigmentaria (RP), abarca todo un grupo de enfermedades genéticas raras, que implican degeneración de células de la retina. Su manifestación es crónica, progresiva, bilateral y de aparición tardía (infancia o adolescencia). Etimológicamente, el término retinosis recoge afectación de la retina, que está inflamada, mientras que el término pigmentaria, hace referencia a la acumulación de pigmento en el ojo. La degeneración inicial afecta principalmente a bastones, produciendo ceguera nocturna y visión en túnel, ya que los bastones son los responsables de la visión en condiciones de baja luminosidad y su distribución es principalmente periférica. En fases avanzadas, la enfermedad afecta a los conos, por lo que se ve comprometida la visión de colores, produciéndose finalmente ceguera.

En este grupo de enfermedades de origen genético, la convergencia fenotípica se puede alcanzar a partir de multitud de mutaciones. Hasta la fecha, se han descrito gran número de mutaciones asociadas con RP, el 20-25% de ellas son autosómicas dominantes, un 15-20% autosómicas recesivas y un 10-15% son ligadas al cromosoma X. Según el modo de afectación, podemos distinguir entre variantes sindrómicas (20-30%), como el síndrome de Usher o el síndrome de Bardet-Biedl, o bien limitadas a los ojos. Muchas de estas mutaciones se podrán expresar tanto en fotorreceptores como en células del epitelio pigmentario de la retina (1).

Las células fotorreceptoras (conos y bastones) son en realidad un tipo de neuronas altamente especializadas en la transformación del estímulo luminoso en impulsos eléctricos, la fototransducción. Tienen una organización altamente polarizada, con un segmento principal donde reside la maquinaria molecular responsable de los procesos de biosíntesis, de obtención de energía y de transporte a través de membrana y un segmento exterior constituido por discos membranosos rodeados de membrana plasmática, donde, alrededor del 90% del contenido proteico es rodopsina.



No obstante, la labor de los conos y bastones requiere de la actividad de las células del epitelio pigmentario de la retina. Las células RPE (*Retinal Pigment Epithelium*), constituyen una monocapa de células pigmentadas que conforman buena parte de la barrera hemato-retinal. Son células polarizadas que en la región apical, presentan largas microvellosidades que interdigitan con el segmento externo de las células fotorreceptoras, mientras que en la porción basolateral se une a la membrana de Bruch, que la separa de la capa coriocapilar. En sentido estricto, son células auxiliares, que desempeñan gran número de funciones, de las que depende la supervivencia de las células fotorreceptoras. Estas células son responsables de la nutrición de conos y bastones, además del mantenimiento de la homeostasis, dado que controlan la composición del espacio subretinal. También están implicadas en la fagocitosis y reciclado de componentes (re-isomerización del 11-cis-retinal), de las porciones distales del segmento externo (2). Además, secretan inmunomoduladores que convierten a la retina en un espacio inmunológicamente privilegiado (3). En consecuencia, de lo anteriormente expuesto cabe inferir, que desordenes que afectan a las células RPE, automáticamente se traducen en degeneración de los fotorreceptores.

Muchas de las mutaciones subyacentes en RP, están relacionadas con factores de *splicing*, como PRPF31, PRPF3 o PRPF8 (4–6). En este estudio nos centraremos en la valoración de la RP, en el contexto de la mutación V2325fsX2329 asociada a PRPF8 (4).

El *splicing* alternativo (AS) constituye un nivel de regulación de la función génica importante en eucariotas. La mayor parte de los genes humanos están sometidos a AS de forma que se aumenta la complejidad y diversidad del proteoma (7). Desde la perspectiva del producto génico, el AS puede implicar un cambio en la localización subcelular o incluso hacer que sea secretado, puede variar las propiedades enzimáticas o cambiar la forma en que interacciona con otras proteínas (8). A nivel tisular, se ha descrito que el perfil de AS está muy conservado en tejidos complejos, altamente especializados,

como el tejido nervioso (9). Además, juega un papel importante en diversos procesos como la diferenciación tisular, la división o la senescencia celular (10). El AS es en definitiva un proceso bien regulado que se integra junto con otros mecanismos en la regulación de la expresión génica, cuya alteración se puede relacionar con multitud de enfermedades y con la modificación de las propiedades celulares (8). Tomando esto en cuenta, no ha de extrañar que este sea el mecanismo subyacente en muchos casos de RP.

En humanos, dada la imposibilidad de obtener y cultivar células fotorreceptoras, las estrategias de análisis de los mecanismos subyacentes a la RP, se centran en el estudio de células RPE. En este sentido, existen protocolos para la obtención de *inducted pluripontential Stem Cells* (IPSC), generadas a partir de fibroblastos y después diferenciar éstas hacia células a RPE (11,12).

Una herramienta útil para la valoración integral de enfermedades, es la secuenciación de alta resolución, conocida como *Next Generation Sequencing* (NGS). Las técnicas NGS comparten una serie de pasos independientemente de la plataforma, que comprenden, en primer lugar, una fragmentación del DNA/cDNA de partida, a continuación, una amplificación clonal de los fragmentos generados y finalmente, la secuenciación y en paralelo de los distintos fragmentos. De esta forma se generan multitud de lecturas cortas, que han de ensamblarse para recomponer el la secuencia original.

De todo ello surge un desafío importante para la bioinformática, el manejo de grandes cantidades de información. A partir de aquí, las técnicas NGS pueden tener diversas aplicaciones, como la detección de variantes genómicas, el análisis del *copy number variation* (CNV), el estudio epigenético, el análisis de expresión y *splicing* alternativo, etc.

Uno de los ejemplos más populares de plataforma NGS, lo constituye la plataforma Illumina (Solexa), cuya estrategia está esquematizada en la figura 1. Aquí la fragmentación del DNA se realiza por nebulización, tras lo cual se lleva a cabo la reparación de extremos y ligamiento enzimático de dos adaptadores distintos P5 y P7.

La fase de amplificación se realiza en fase sólida, en la llamada *flowcell*. Los fragmentos de DNA con los adaptadores, se unen a una superficie sembrada de oligos complementarios, que actuarán como *primers* en el proceso de amplificación en puente por PCR. A continuación, los moldes originales son eliminados, dejando únicamente las copias unidas covalentemente al sustrato. Tras la amplificación, se generan del orden de 1000 copias por fragmento. En el proceso se forman dos *clusters* complementarios de los segmentos de DNA, atendiendo al adaptador por el que se encuentren covalentemente unidos a la *flowcell*. Después, se elimina una de las cadenas actuando sobre uno de los dos tipos de adaptadores, para asegurar que la lectura al secuenciar sea en la misma dirección.

Para la fase de secuenciación, al igual que en la secuenciación Sanger, es necesario incorporar *primers* sobre los que ir adicionando los nucleótidos. Tras los *primers*, se añaden nucleótidos marcados con fluorocromos, que al incorporarse conforme a complementariedad, emiten una fluorescencia específica de nucleótido, que es detectada por el circuito fotográfico. Tras esto, se lavan las bases no incorporadas y se repite el ciclo. En consecuencia, en cada ciclo se asigna un nucleótido a cada *cluster* y por repetición del proceso se obtiene la secuencia.

A partir de aquí, se pueden plantear dos modalidades, single-end o pair-end, en función de si el proceso finaliza aquí y nos quedamos solo con una de las lecturas (single-end) o bien se repite el proceso, eliminando esta vez los fragmentos anclados covalentemente por medio del otro adaptador (pair-end).

Dentro del contexto de la secuenciación masiva, la tecnología RNAseq (RNA *sequencing*), permite cuantificar las proporciones relativas en un determinado momento de distintas poblaciones de RNA, tales como mRNA, miRNA, rRNA, siRNA, tRNA, etc (13,14). Por tanto, puede emplearse para valorar patrones de expresión diferencial y análisis de *splicing*, al seleccionar el *pool* de mRNA. En todos los casos, se debe preparar una librería, lo cual requiere la extracción de RNA, seguida de la selección de la población de RNA ,sobre la cual se va a construir la librería, para finalmente sintetizar el cDNA. Para la valoración de los mRNA, se seleccionan los RNAs poliadenilados.

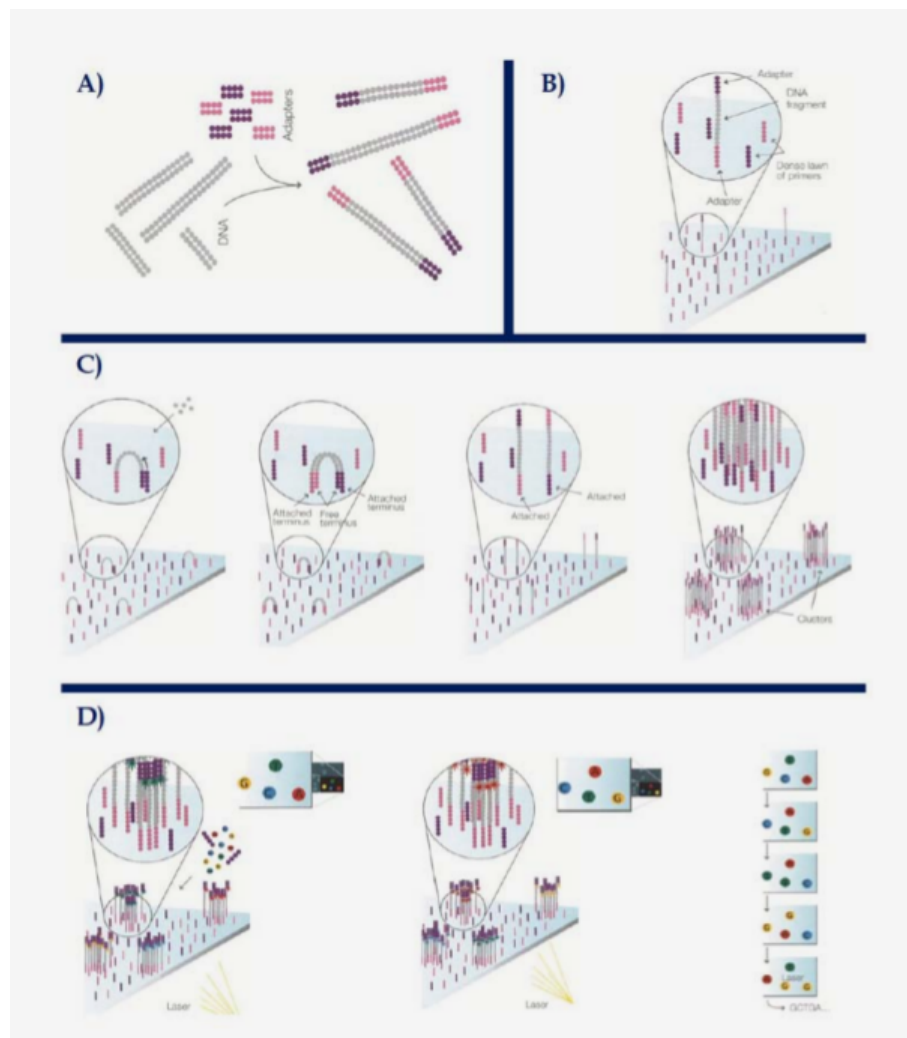


Fig. 1: Producción de clusters por amplificación en puente empleada por illumina en la secuenciación por síntesis.

En RNAseq el desafío radica en la identificación y cuantificación de los transcritos a partir de los reads obtenidos en la secuenciación. En organismos donde no hay distintos transcritos por gen, como los procariotas, la tarea resulta más sencilla, dado que no hay ambigüedad en el alineamiento. Sin embargo en eucariotas, debido al *splicing* alternativo, la cuantificación se complica considerablemente. Para abarcar el problema, los modelos han debido evolucionar, como también lo ha hecho la técnica. A este respecto los modelos que mapean frente a transcriptoma de referencia plantean un sesgo asociado, mayor o menor en función del grado de conocimiento del transcriptoma en cuestión (15). No obstante, en organismos para los que la anotación de isoformas es bastante completa, como en humanos, este sesgo es considerablemente bajo.

La creciente información de isoformas aportada por el uso de estas nuevas tecnologías es recogida en bases de datos, como APPRIS (16). APPRIS es una base de datos desarrollada por el Centro Nacional de Investigaciones Oncológicas (CNIO), dentro del consorcio GENCODE, donde se relacionan las distintas variables descritas para un determinado gen, con diversas características biológicamente relevantes, para seis especies de vertebrados (humano, ratón, rata, cerdo, chimpancé y pez cebra) y dos de invertebrados (*Caenorhabditis elegans* y la mosca de la fruta). APPRIS proporciona anotaciones manuales de isoformas, a partir de datos estructurales y funcionales de proteínas. Para ello, toma como entrada las secuencias de transcritos y polipéptidos e integra la información estructural y funcional al respecto, de 8 webs (firestar, Matador3D, SPADE, CORSAIR, INERTIA, CExonic, THUMP y CRASH) para realizar las anotaciones y clasificar las isoformas principales (16).

## 4. MATERIALES Y MÉTODOS

### 4.1 MATERIAL BIOLÓGICO

El material biológico del que partimos, es el RNA extraído de 12 muestras distintas, 6 muestras de un paciente con retinosis pigmentaria (3 de fibroblastos y 3 de células RPE) y 6 muestras de control sano (3 de fibroblastos y 3 de células RPE). Las células RPE, se obtuvieron a partir de IPSC (*Inducted Pluripotencial Stem Cels*). Las IPSC, son células madre pluripotenciales inducidas, obtenidas a partir de células adultas que inicialmente no lo son, mediante un protocolo determinado (12). El paciente de retinosis pigmentaria presenta la mutación V2325fsX2329 en el gen PRPF8, en heterocigosis, determinada por una delección de 21bp, que codifica para un Factor de *splicing*, figura 2. Esta mutación se asocia a casos de retinosis pigmentaria (4).

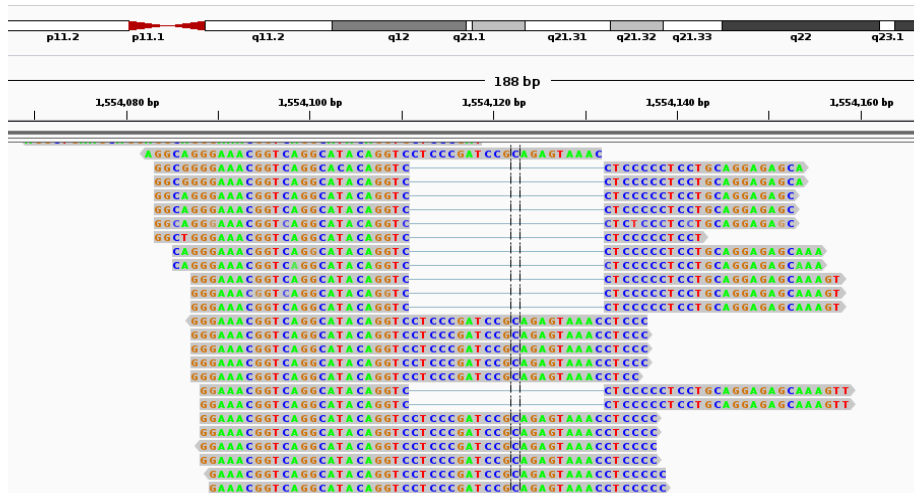


Fig. 2: Visualización de lecturas de RNA-seq en IGV, procedentes de las células RPE del paciente.

Los datos de RNA-seq se obtuvieron por el protocolo de secuenciación masiva *single-end*, en la plataforma de Illumina HiSeq 2500.

## 4.2. PIPELINE Y SOFTWARE EMPLEADO

Los datos de RNAseq se recibieron del departamento de genómica en formato fastq. Para el análisis de dichos datos, procedimos conforme a la *pipeline* representada en la figura 3.

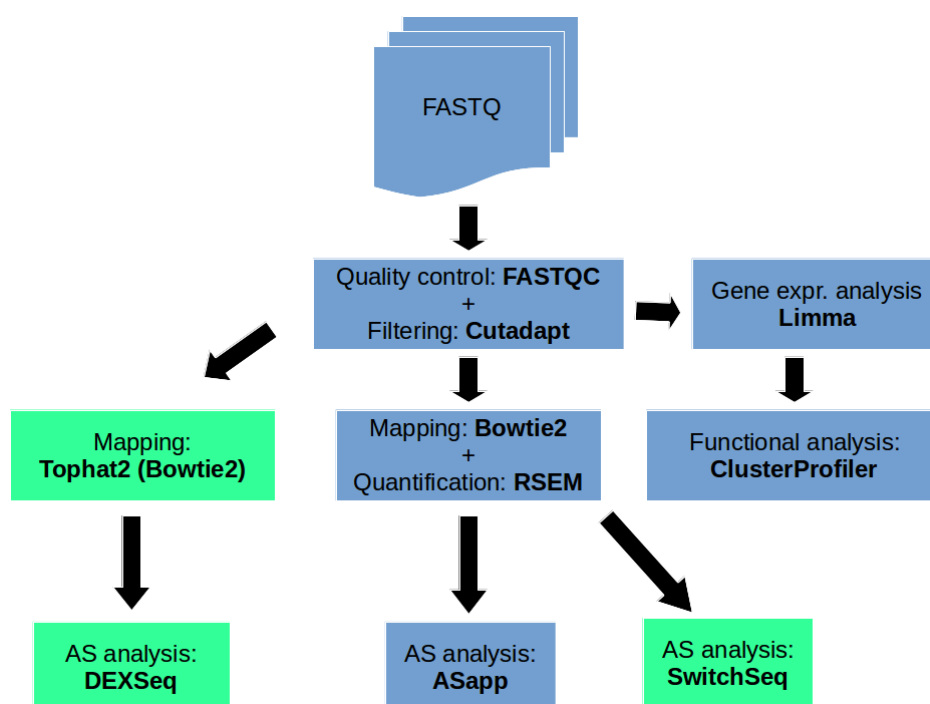


Fig. 3: Pipeline seguida para el análisis de los datos RNAseq.

En primer lugar, comprobamos la calidad de las lecturas mediante FastQC v11.5 (17), en busca de sesgos en la composición de nucleótidos, presencia de adaptadores, regiones de baja calidad en las lecturas, etc.

A continuación, se procede a la eliminación de adaptadores y filtrado de reads con poca calidad o longitud insuficiente, con el software cutadapt-1.7.1. (18)).

El alineamiento de las *reads* se llevó a cabo frente a un transcriptoma de referencia, con Bowtie2. Mediante RSEM-3.2.3. se estimó el nivel de

expresión de genes e isoformas, ya que computa la probabilidad de que un determinado fragmento corresponda a un transcrito concreto, mediante el algoritmo EM (Expectation Maximization)(19,20).

Para el análisis de expresión diferencial utilizamos el paquete de Bioconductor Limma-3.32.15(21), en R version 3.4.1. El método voom del paquete limma, realiza la transformación de nuestros conteos discretos a datos continuos, dado que limma es una herramienta potente adaptada a datos de microarrays (22). El paquete limma ajusta, en primera instancia, un modelo lineal a cada gen, aplicando *empirical Bayes* para moderar el error estándar de las estimaciones del  $\log_2(\text{fold-change})$ , a partir de información tomada del conjunto general de genes(23).

El análisis funcional se orientó a familias de genes representantes de procesos biológicos de Gene Ontology, rutas metabólicas de KEGG y enfermedades de Disease Ontology(24–26). Para ello, se hizo un estudio de sobrerrepresentación (ORA) o bien en su caso, un análisis de tipo GSEA (Gene Set Enrichment Analysis). Esto se hizo empleando el paquete de Bioconductor clusterProfiler v3.4.4(27), representando los resultados mediante el paquete igraph v1.1.2 y el programa gsea2-2.2.0(28).

Adicionalmente, construimos una firma de genes (gene sets), valiéndonos de herramientas de *text-mining* (Génie web server) con genes de interés en retinosis pigmentaria(29). Por este procedimiento obtuvimos una colección de 297 para un  $\text{FDR} < 0.001$ . A esta lista de genes adicionamos 21 genes aportados por expertos en RP, de los cuales 15 ya estaban en la lista obtenida por *text mining*. El *gene set* resultó de la unión de ambas listas de genes, conteniendo un total de 303 genes y fue analizado por medio del programa gsea 2-2.2.0.

Para la detección de eventos de *splicing* alternativo, entre muestras de paciente y control, desarrollamos una herramienta web, ASapp.



Paralelamente, quisimos procesar los datos con otras dos herramientas que abordan el estudio de eventos de *splicing* de manera distinta a nuestra aplicación, como son SwitchSeq y el paquete de R bioconductor, DEXSeq v1.22.0(30,31).

SwitchSeq estudia la utilización diferencial de isoformas entre dos condiciones, y toma como entrada los datos proporcionados por RSEM. En primer lugar, identifica las isoformas mayoritarias para gen y condición y en segundo lugar detecta como *switching*, aquellos casos en los que hay cambio de isoforma mayoritaria entre ambas condiciones(30).

Por su parte, DEXSeq, está basado en la utilización diferencial de exones entre condiciones. La entrada de DEXSeq son archivos bam obtenidos del alineamiento con Tophat, que internamente recurre a Bowtie 2, alineando frente a genoma de referencia, cuantificando las lecturas alineadas con exones(32,33). DEXSeq, asume que el numero de *reads* que alinea con exón sigue una distribución binomial negativa. Posteriormente, adapta un modelo lineal logarítmico para predecir la media de *reads* que alinean con un exón concreto, ajustando el modelo para cada gen, de modo que el algoritmo considera un componente que recoge la parte de la fracción de *reads* que alinea con un exón, que varía entre las dos condiciones(31) .

### 4.3. ASapp

La aplicación desarrollada, identifica los eventos *splicing* alternativo en contrastes *pairwise* y realiza una interpretación biológica de dichos eventos, a partir de información obtenida de APPRIS.

ASapp tiene un interfaz gráfico que permite la utilización de la aplicación de manera sencilla, basado en el paquete Shiny v1.0.3(34).

Como entrada, el programa requiere varios archivos y un directorio:

1. Un fichero con la información de APPRIS, tal como el biotipo (clasificación completa de biotipos en [http://www.gencodegenes.org/gencode\\_biotypes.html](http://www.gencodegenes.org/gencode_biotypes.html)) y el identificador *Consensus Coding Sequence*.
2. Un archivo que contenga una tabla que defina a que condición pertenece cada muestra, a partir del cual se ofrece la posibilidad de elegir la condición de referencia frente a la que comparar la condición problema.
3. Otro archivo que contenga aquellos genes con expresión mínima de 1 cpm (*counts per million*) y la información del p-valor ajustado y *fold change* del análisis de expresión diferencial.
4. Un directorio donde estén los ficheros con los valores de *isoform percentage* para cada gen, proporcionado por RSEM.

De esta forma ASapp realiza un primer filtrado de datos, al tener en cuenta genes con valor de expresión mínima de 1 cpm (*counts per million*), considerando irrelevante lo que suceda en genes con una expresión inferior. Además, ofrece la posibilidad de realizar un segundo filtrado, para quedarnos con los genes que no se expresan diferencialmente en ambas condiciones, de manera significativa.

La detección de genes candidatos para eventos de *switching*, se basa en el cálculo de la distancia de Jensen-Shannon para cada gen entre ambas condiciones. Esta estrategia es seguida por diversas herramientas destinadas a valorar eventos de *switching* (35–37). La raíz cuadrada de la divergencia de Jense-Shannon, se emplea para hacer un análisis cluster que nos permita clasificar los genes en *Switching/No-switching*.

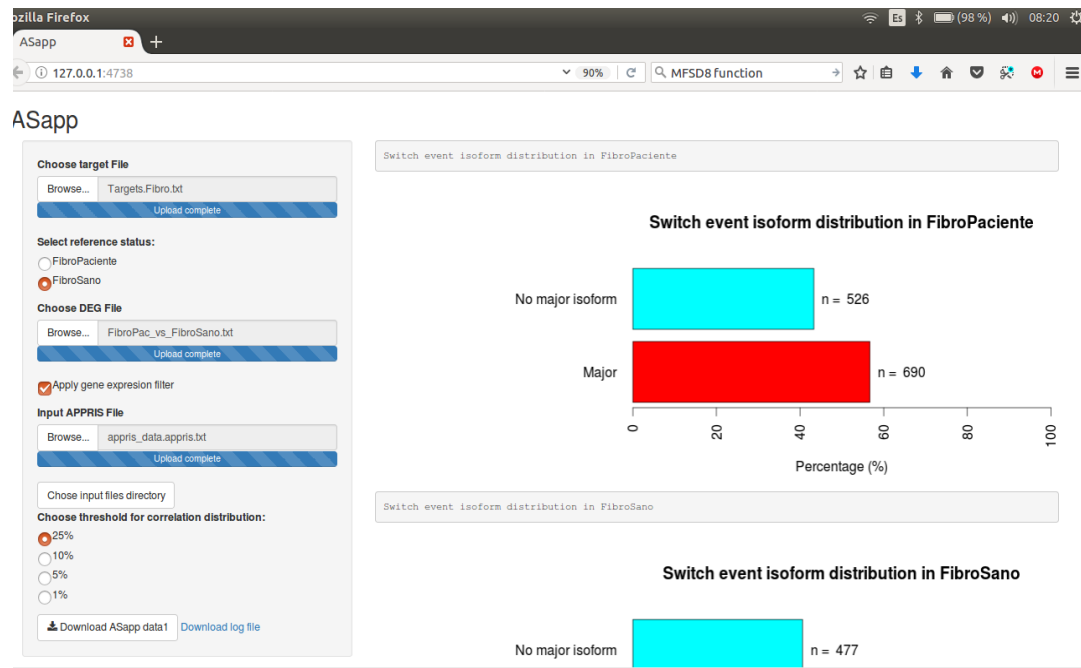
Una vez seleccionados los genes con mayor divergencia, analizamos la correlación entre el fenotipo/condición y el valor promedio del porcentaje de isoforma, a nivel de transcrito. A este nivel, aplicaremos un filtrado adaptado a nuestros datos, cuya astringencia está predefinida por el usuario, pero en ambos sentidos, por arriba y por abajo de nuestra distribución de índices de correlación. De este modo nos quedaremos con los transcritos que mayor correlación presenten para nuestros fenotipos. En nuestro estudio seleccionaremos el 25% de los casos que mayor correlación con el fenotipo presenten.

Tras identificar las isoformas sujetas a *switching*, el programa realiza anotaciones y estadística básica orientada a la interpretación biológica de los fenómenos de *switching* detectados. Además el programa realiza un análisis de sobre-representación de términos GO para procesos biológicos, a partir de los genes que experimentan *switching*.

El software, ofrece la posibilidad de descarga de dos ficheros. Uno de ellos, un log file, que registra los ficheros y parámetros utilizados en el análisis, el otro es una tabla csv con los resultados del análisis.

En la salida html, se representan gráficamente los resultados de distribución de las isoformas implicadas en el *switching* en función de su abundancia y en función del resultado del *splicing*, para cada condición del contraste pairwise. Además, se despliegan los datos resultantes en una o dos tablas interactivas, en función de si el usuario seleccionó el filtro de expresión diferencial de

genes. Por último, se despliega una gráfica de tipo dotplot, con las 20 familias de genes más representadas, referidas a procesos biológicos de *Gene Ontology*.

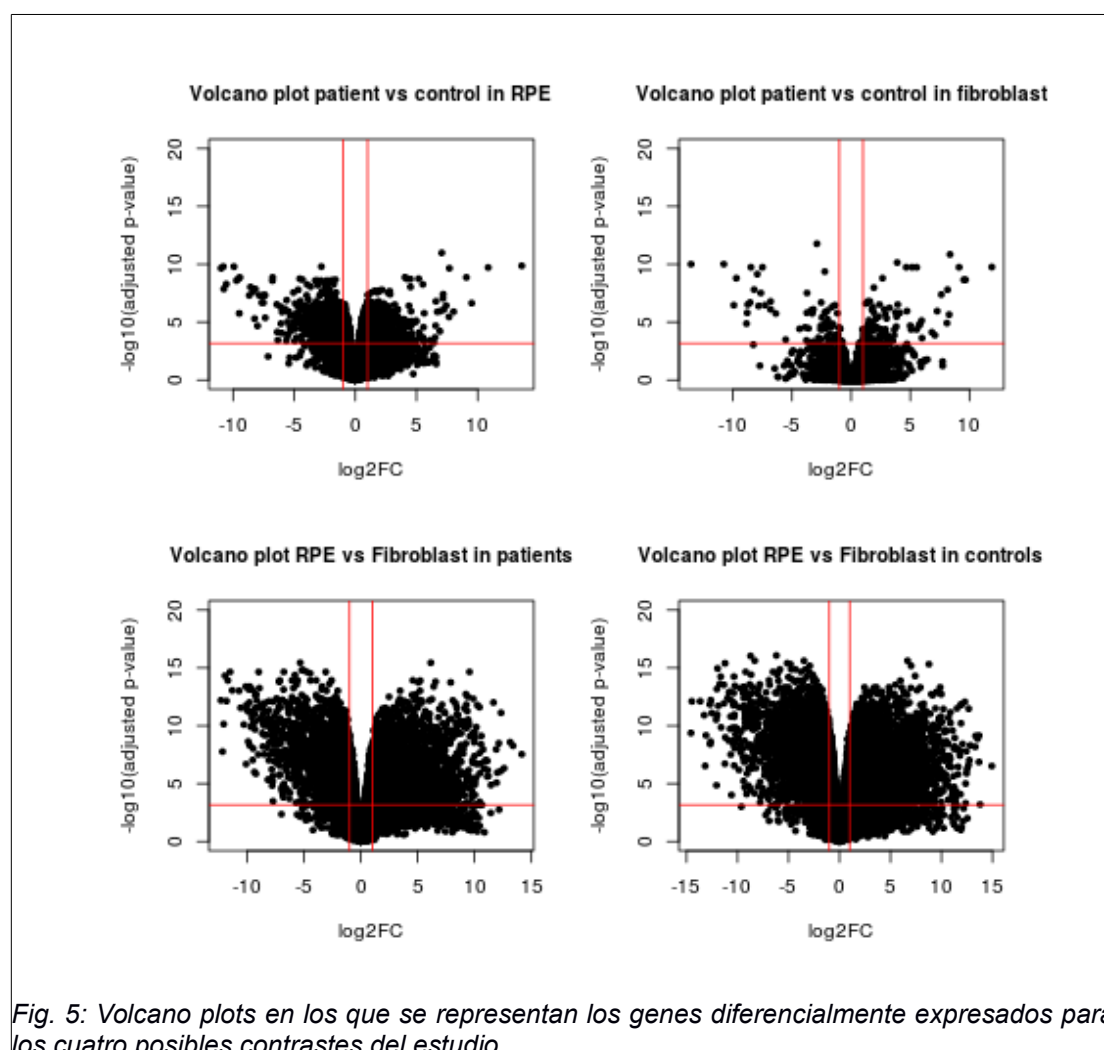


*Fig. 4: Interfaz gráfico del programa ASapp, donde se muestra un panel lateral para la entrada/salida de archivos y un panel principal para gráficas y tablas que recogen los resultados.*

## 5. RESULTADOS

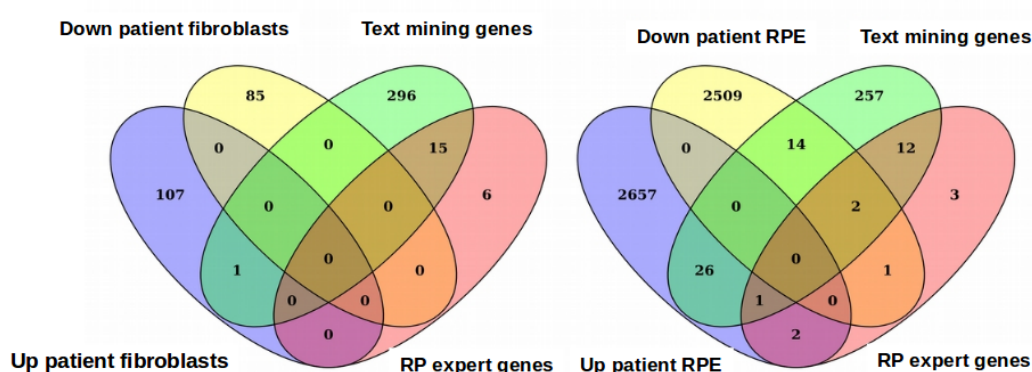
### 5.1 ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE GENES

El número de genes expresados diferencialmente fue mayor al comparar fibroblastos y RPE, tanto en pacientes como en controles, que al orientar el análisis en función de la presencia/ausencia de RP. Al valorar la expresión diferencial entre muestras de paciente y control, para un mismo tipo celular, la cantidad de genes que se expresan diferencialmente fue mayor en células RPE que en fibroblastos, figura 5.



*Fig. 5: Volcano plots en los que se representan los genes diferencialmente expresados para los cuatro posibles contrastes del estudio.*

Como se expuso en el apartado de materiales y métodos, uno de los abordajes planteados para este estudio, fue analizar el comportamiento de los genes de la lista RP (lista de 303 genes relacionados con RP). En células RPE, al comparar lo observado entre paciente y control, se encontraron 61 genes de la lista RP diferencialmente expresados, figura 6. En fibroblastos, solo el gen RIMS1 se encontró diferencialmente expresado (sobre-expresado en paciente).



*Fig. 6: Diagramas de Venn que representan la proporción de genes de las lista de genes relevantes en retinosis pigmentaria en relación a los genes diferencialmente expresados en células RPE (izda) y fibroblastos (dcha).*

En las tablas 1 y 2 se muestran con detalle los genes de la lista RP, que están activados y reprimidos de forma significativa, para un p-valor ajustado inferior a 0.05 (False Discovery Rate, ajustado mediante el método de Benjamini-Hotchberg).

Table 1. DEG Upregulated in patient RPE .				
ENSEMBL	Symbol	Adj. p.value	Fold change	Source
ENSG00000166278	C2	0.003808518	6.544212	Text mining genes
ENSG00000100997	ABHD12	0.003994548	2.250226	Text mining genes
ENSG00000119681	LTBP2	0.004001815	8.533637	Text mining genes
ENSG00000243649	CFB	0.004301115	4.674077	Text mining genes
ENSG00000156973	PDE6D	0.005901446	2.048458	Text mining genes
ENSG00000108370	RGS9	0.006856765	3.086165	Text mining genes
ENSG00000198513	ATL1	0.006891962	3.671589	Text mining genes
ENSG00000136425	CIB2	0.00888479	3.058602	Text mining genes
ENSG00000065154	OAT	0.00968424	1.615117	Text mining genes
ENSG00000118217	ATF6	0.01062804	1.431444	Text mining genes
ENSG00000197579	TOPORS	0.01169506	1.429496	Text mining genes
ENSG00000125741	OPA3	0.01312338	1.366286	Text mining genes
ENSG00000122550	KLHL7	0.01509543	1.35235	Text mining genes
ENSG00000172071	EIF2AK3	0.01727047	1.532755	Text mining genes
ENSG00000124659	TBCC	0.01735226	1.76572	Text mining genes
ENSG00000091482	SMPX	0.02544323	2.462372	Text mining genes
ENSG00000171202	TMEM126A	0.02687646	1.607885	Text mining genes
ENSG00000186868	MAPT	0.02851929	5.17143	Text mining genes
ENSG00000163359	COL6A3	0.02954519	15.23836	Text mining genes
ENSG00000123240	OPTN	0.03068763	1.665258	Text mining genes
ENSG00000135338	LCA5	0.03118708	1.570401	Text mining genes
ENSG00000102119	EMD	0.03611101	1.388789	Text mining genes
ENSG00000122507	BBS9	0.03902125	1.343225	Text mining genes
ENSG00000118402	ELOVL4	0.04019646	1.915107	Text mining genes
ENSG00000215612	HMX1	0.04633148	18.51319	Text mining genes
ENSG00000138175	ARL3	0.01409435	1.961308	Text mining & expert genes
ENSG00000102931	ARL2BP	0.009002576	1.741651	Expert genes
ENSG00000176842	IRX5	0.03428439	27.25287	Expert genes

Además de los genes C2 y CFB, que codifican para integrantes del sistema de complemento, se encontraron activados genes implicados en la respuesta inflamatoria como IL6, las quimioquinas, CXCL1 ,CXCL6, CXCL3, CXCL12 y CXCL16, el inhibidor del complejo de ataque, CD59 o la proteína de complemento C1QL1, para un valor de FDR<0.05.

Table 2. DEG downregulated in patient RPE.				
ENSEMBL	Symbol	Adj. p.value	Fold change	Source
ENSG00000162769	FLVCR1	0.002168256	-2.215566	Text mining genes
ENSG00000144061	NPHP1	0.008660831	-2.148958	Text mining genes
ENSG00000113971	NPHP3	0.01004924	-1.986763	Text mining genes
ENSG00000177082	WDR73	0.01004924	-1.727773	Text mining genes
ENSG00000157796	WDR19	0.01090329	-1.899932	Text mining genes
ENSG00000095397	WHRN	0.01372455	-1.614969	Text mining genes
ENSG00000042317	SPATA7	0.01372455	-1.614969	Text mining genes
ENSG00000079841	RIMS1	0.01488952	-2.157239	Text mining genes
ENSG00000096093	EFHC1	0.02750142	-2.053264	Text mining genes
ENSG00000164073	MFSD8	0.02831847	-1.558001	Text mining genes
ENSG00000117360	PRPF3	0.02914364	-1.409375	Text mining genes
ENSG00000138002	IFT172	0.03127648	-1.580643	Text mining genes
ENSG00000131697	NPHP4	0.0397236	-1.426432	Text mining genes
ENSG00000198707	CEP290	0.04828991	-1.57974	Text mining genes
ENSG00000046651	OFD1	0.004687297	-1.738562	Text mining & expert genes
ENSG00000167434	CA4	0.04370844	-16.5627	Text mining & expert genes
ENSG00000112232	KHDRBS2	0.0006483311	-174.7232	Expert genes

En fibroblastos, tan solo el gen RIMS1 (ENSG00000079841), de los genes RP presentó expresión diferencial entre pacientes y controles (adj. p.value=0.002920719, fold change=13.49325).

## 5.2. ANÁLISIS FUNCIONAL

### 5.2.1. ANÁLISIS FUNCIONAL EN RPE

El análisis ORA (*Over Representation Analysis*) para grupos de genes que integran las rutas de KEGG, se realizó tomando un p-valor ajustado por el método de Benjamini-Hotchberg menor de 0.05. Los resultados se muestran en la figura 7. Como cabría esperar, dada la mutación del paciente, están sobrerrepresentados genes involucrados en la formación del spliceosoma . Más relevante, resulta la sobrerrepresentación de familias de genes que participan en el procesamiento proteico en el retículo endoplasmático, la exportación de proteínas, la fosforilación oxidativa y la formación de proteosomas. Todos los integrantes del proteosoma se encontraron sobre-



expresados en RPE de paciente. También, resultó significativa la representación de rutas de genes relacionados con el desarrollo de enfermedades neurodegenerativas como el Parkinson o el Alzheimer, rutas implicadas en la respuesta a mononucleosis infecciosa, o bien procesos como la meiosis o la recombinación de homólogos.

Dentro de la familia de genes que participan en el procesamiento de proteínas en el RE, 59 de los 63 genes pertenecientes a este grupo estaban sobre-expresados en RPE de paciente, incluyendo todos los genes efectores que conforman el sistema UPR. Adicionalmente, el XBP1 (no incluido en la familia de genes), regulador positivo de UPR, estaba diferencialmente (Fold Change=2.19, adj.p-value=0.004) activado en el paciente.

En el grupo de genes participantes en la ruta de la fosforilación oxidativa, 50 de los 58 genes diferencialmente expresados, estaban activados en el paciente.

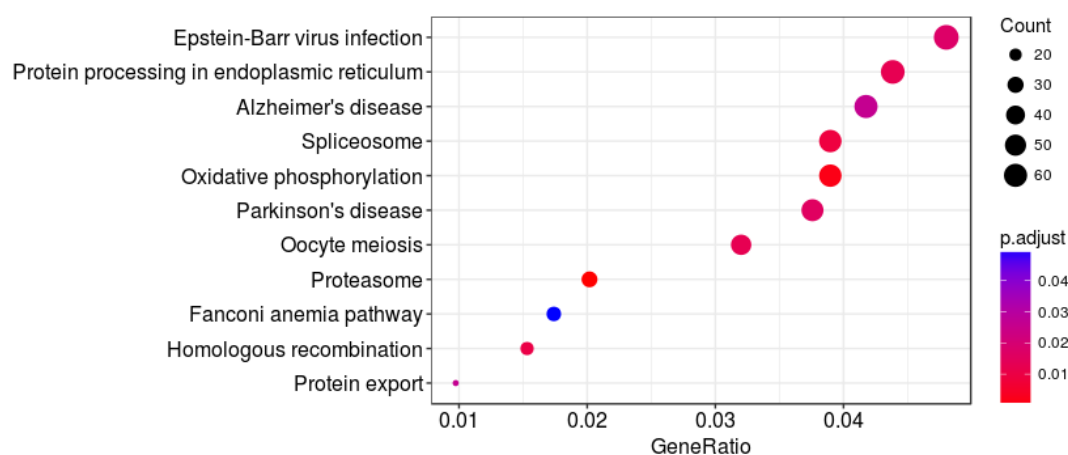


Fig. 7: Resultados del análisis de sobrerepresentación para rutas de KEGG en genes diferencialmente expresados entre RPE paciente y RPE control (FDR<0.05).

En la figura 8, están representados los grupos de genes sobrerepresentados, para procesos biológicos de Gene Ontology, en el análisis ORA.



*Fig. 8: Red de interacciones entre los distintos procesos biológicos de Gene Ontology obtenidos en el análisis de sobrerrepresentación en genes diferencialmente expresados entre RPE paciente y RPE control (FDR<0.01).*

En la figura 9, están representados los 20 *gene sets* más significativos para procesos biológicos de GO. En total, 30 gene-sets resultaron estadísticamente significativos para un punto de corte de q-value de 0.001. En consonancia con lo observado para *pathways* de KEGG, entre los procesos biológicos más significativos, destacan aquellos relacionados con el *splicing* y con la degradación/procesamiento de proteínas por ubiquitinización. Entre los genes diferencialmente expresados (activados en RPE de paciente) relacionados con la ubiquitinización están NEDD4, TOPORS, SIAH1, KLHL7, RanBP2, PIAS3, UCH-L3, UCH-L1/PGP9.5, USP11, UBE2B, UBE2K y UBE2E3.

Las dos principales consecuencias de la activación del sistema UPR, son la activación de UPS y la detención del ciclo celular, lo cual concuerda con la sobrerrepresentación de la familia de genes que participa en la detención del ciclo celular.

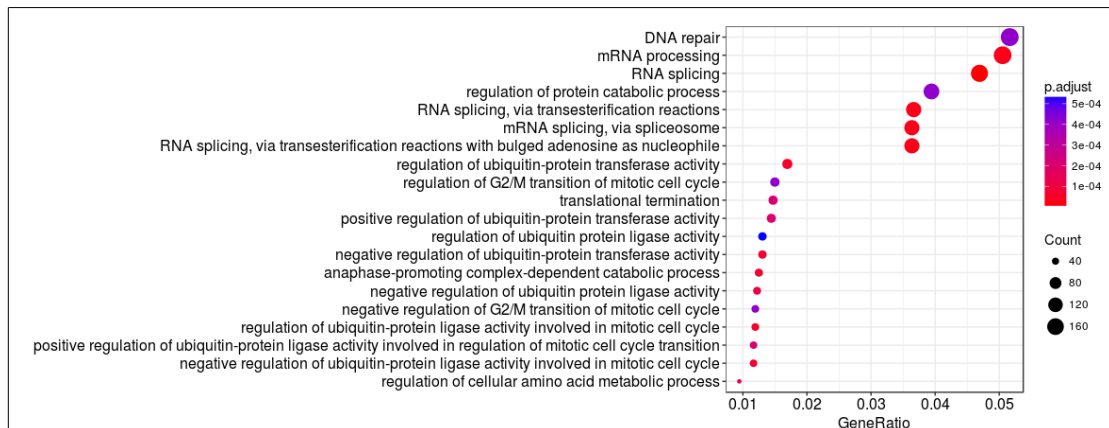


Fig. 9: Resultados del análisis de sobrerepresentación para procesos biológicos de Gene Ontology en genes diferencialmente expresados entre RPE paciente y RPE control (FDR<0.01).

El análisis GSEA para firmas de genes de Disease Ontology, entre otras enfermedades, mostró una alta representación de genes relacionados con enfermedades de la visión, ojos y estructuras anejas (figura 10). Para este test, se consideró un punto de corte del p-valor ajustado de 0.05.

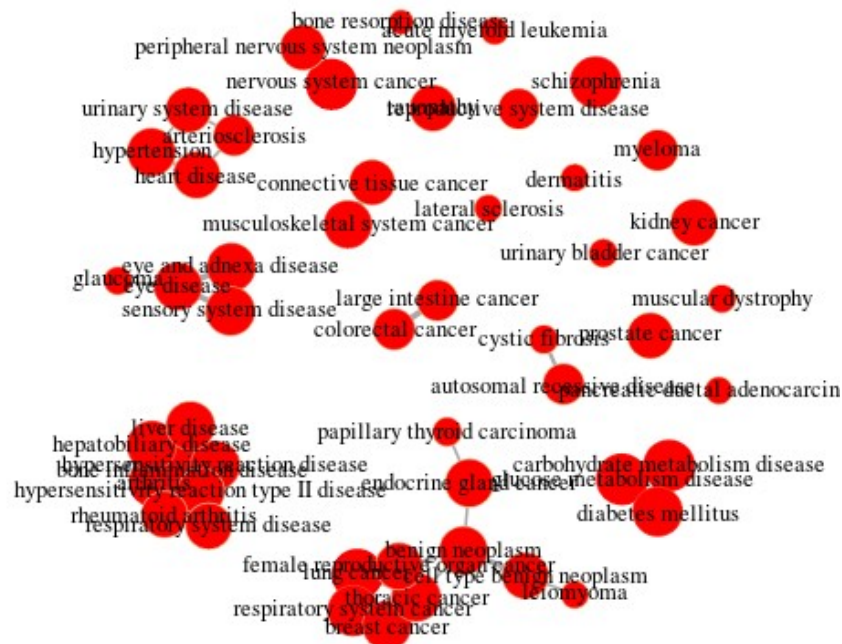


Fig. 10: Red que relaciona los gene-sets que resultaron significativos en el análisis GSEA para términos de Disease Ontology, (FDR<0.05).

Dentro de este grupo de *gene-sets*, seleccionamos, por su especial relevancia en relación con la retinosis pigmentaria, la familia de genes implicados en la degeneración de la retina (DOID:8466), figura 11.

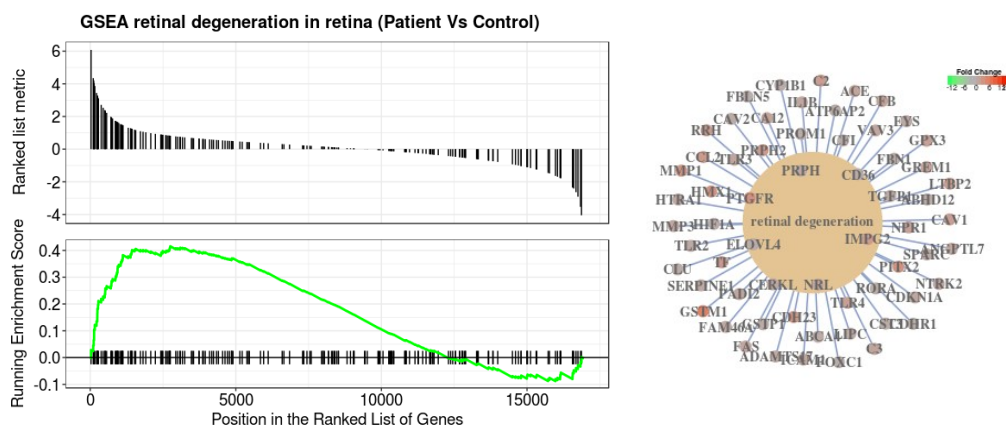


Fig. 11: Gráfica que representa el Enrichment score para la familia de genes implicada en la degeneración de la retina (izda) y figura que asocia los genes diferencialmente expresados en el gene-set de degeneración de la retina con el valor de fold-change (dcha).

En el análisis GSEA se obtuvo un NES=1.736359 (ES=0.42, adj. p.value = 0.004262289, q-value = 0.001209672, set size=203). En la figura 11, a la derecha, se muestran los genes del grupo, coloreados conforme al *fold change* y como puede apreciarse, la inmensa mayoría se encontrarían sobre-expresados. HMX1, LTBP2, C2, CFB, ABHD12, ELOVL4 son genes de la lista RP, que se incluyen en esta familia de genes y están sobre-expresados de forma significativa.

**En resumen, la fosforilación oxidativa, el sistema UPR y la vía UPS de degradación de proteínas se encuentran activados en RPE de paciente en relación con RPE de control.**

### **5.2.2. ANÁLISIS FUNCIONAL EN FIBROBLASTOS**

En fibroblastos, no se obtuvo ningún resultado significativo dentro del análisis funcional de expresión diferencial entre paciente y control.

### **5.3. RESULTADOS DEL ANÁLISIS DE *SPLICING* ALTERNATIVO**

Como ya se expuso en el apartado de material y métodos, la valoración de la presencia de *splicing* alternativo se hizo a través de tres herramientas distintas: ASapp, SwitchSeq y DEXSeq.

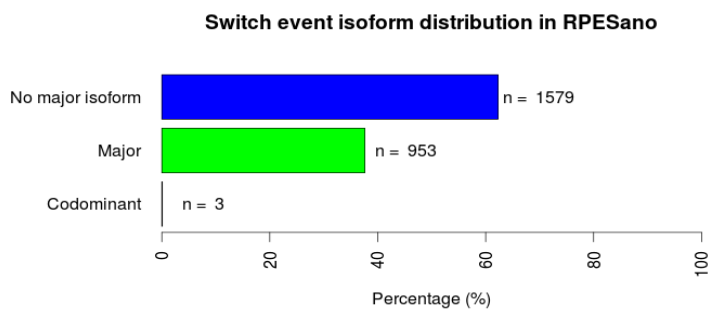
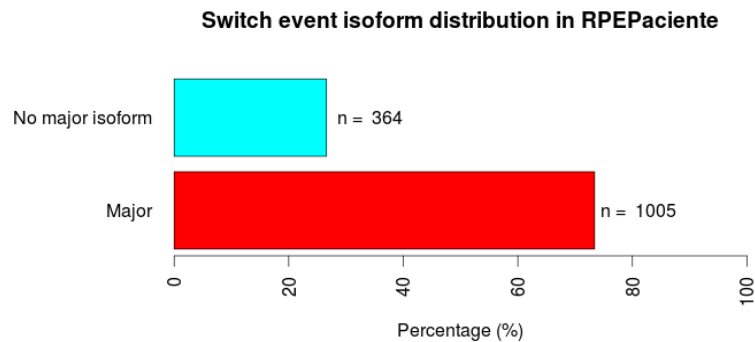
#### **5.3.1. RESULTADOS DE ASapp**

La aplicación ASapp, una vez identifica las isoformas que sufren switching entre las dos condiciones evaluadas, las clasifica conforme a su abundancia relativa dentro de dicha condición y las clasifica según biotipos.

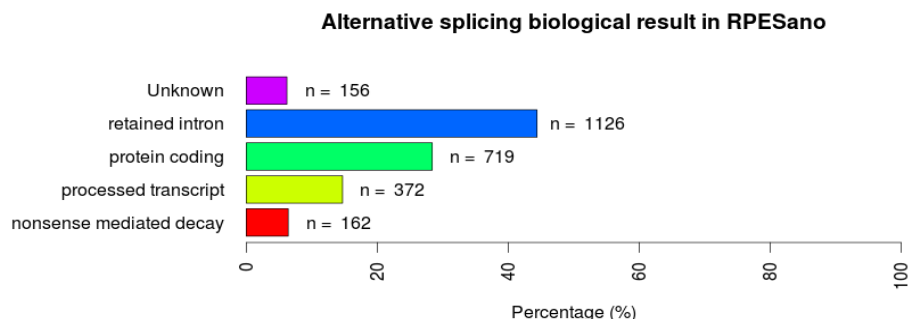
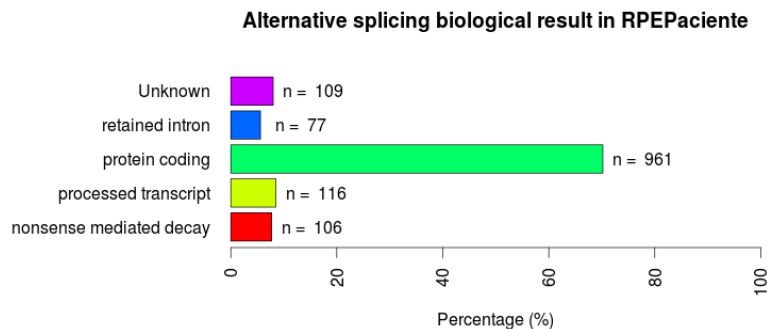
##### **5.3.1.1. Comparación paciente Vs control en células RPE.**

El análisis ASapp en células RPE, entre muestras de paciente y muestras de control, identificó 3904 isoformas sujetas a *splicing* alternativo, de las cuales 2717 (el 70%), correspondían a genes sin expresión diferencial entre condiciones. De ese total, 1366 isoformas estaban asociadas a muestras RPE de pacientes y 2530 asociadas a muestras RPE de control, figura 12 y 13.





*Fig. 12: Distribución de variantes transcripcionales según abundancia relativa en células RPE de paciente(arriba) y en células RPE de control (abajo), en el contraste RPE paciente Vs RPE control.*



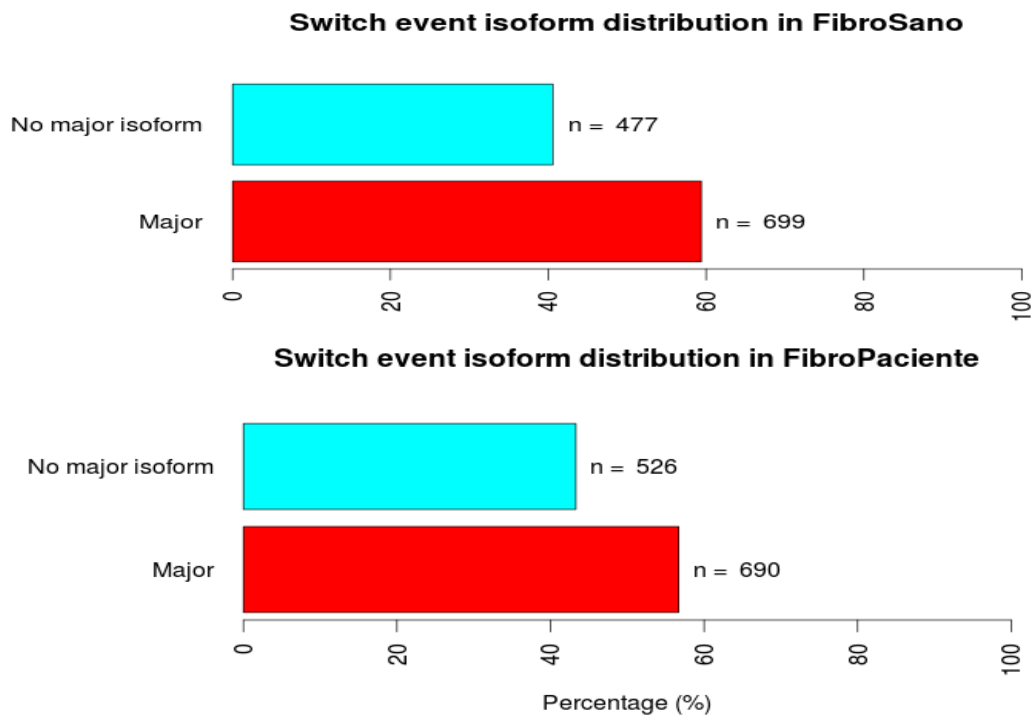
*Fig. 13: Clasificación por biotipos de los transcritos que presentan switching que se asocian con la condición RPE de paciente (arriba) y con la condición RPE de control (abajo), en el contraste RPE paciente Vs RPE control.*

***En este contraste, se observa un patrón de splicing alterado en el paciente, a consecuencia de la mutación, con variantes transcripcionales susceptibles de ser traducidas que aparecen y otras que desaparecen y una disminución del número de isoformas en pacientes, principalmente de isoformas con retención de intrones, respecto del control.***

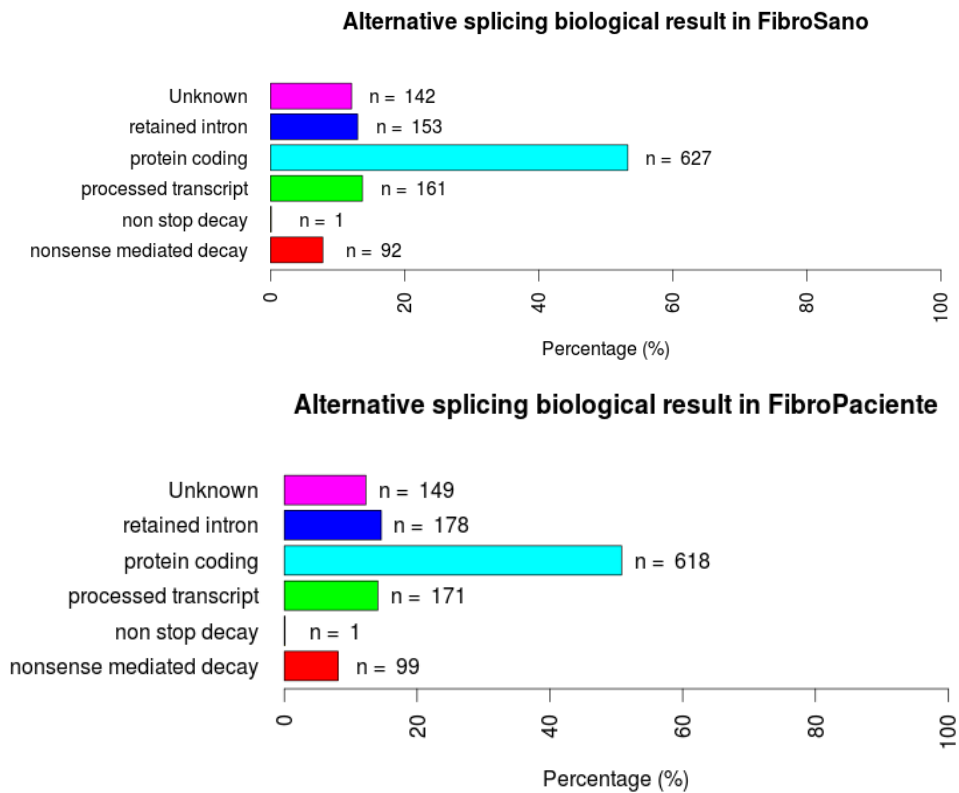
#### **5.3.1.2. Comparación paciente Vs control en fibroblastos**

En el contraste llevado a cabo en fibroblastos, entre muestras de paciente y muestras control, se vieron 2392 isoformas en total sujetas a *switching*, de las que 2259 (94%) resultaron ser de genes no diferencialmente expresados. Del total, se identificaron 1216 isoformas asociadas a muestras de paciente y 1176 asociadas a muestras de control, figuras 14 y 15.





*Fig. 14: Distribución de variantes transcripcionales según abundancia relativa en células fibroblastos de control (arriba) y fibroblastos de paciente (abajo), en el contraste fibroblastos paciente Vs fibroblastos control.*

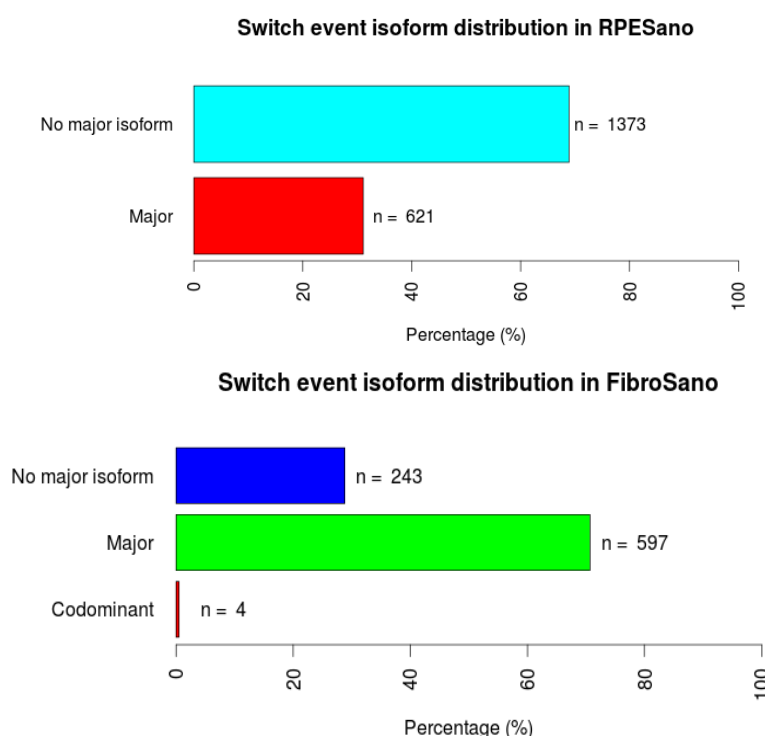


*Fig. 15: Clasificación por biotipos de los transcritos que presentan switching que se asocian con la condición fibroblastos de control (arriba) y fibroblastos de paciente (abajo), en el contraste fibroblastos de paciente Vs fibroblastos de control.*

***En fibroblastos, también hay un buen número de transcritos que presentan switching entre paciente y control, por tanto también está alterado a consecuencia de la mutación. No obstante, en ambas condiciones se da el mismo perfil de splicing, no observándose la marcada reducción de isoformas con retención de intrones, que se veía en el contraste anterior.***

### 5.3.1.3. Comparación fibroblasto Vs RPE en control

En muestras de control, al comparar el uso de isoformas entre fibroblastos y células RPE, se detectaron 2838 variantes transcripcionales cuyas proporciones relativas cambiaron en función del tipo de células, de las cuales 701 (25%) no estaban sujetas a expresión diferencial. En la figura 16, se muestra la clasificación de transcritos que experimentan *switching* en muestras de control, según implicación de la isoforma mayoritaria.



*Fig. 16: Distribución de variantes transcripcionales según abundancia relativa en células RPE de control (arriba) y en fibroblastos de control (abajo), en el contraste fibroblastos control Vs RPE de control.*

En la figuras 17 y 18, se muestra la distribución por biotipos de las variantes transcripcionales, que experimentaron AS conforme al tipo celular en el control.

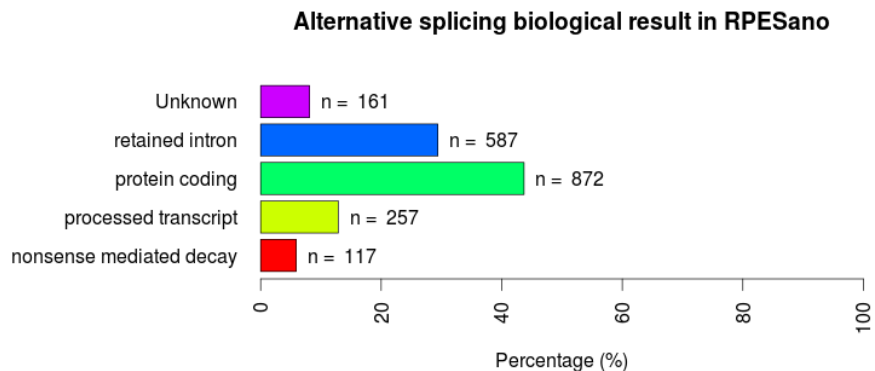


Fig. 17: Clasificación por biotipos de los transcritos que presentan switching que se asocian con la condición células RPE de control, en el contraste RPE de control Vs fibroblastos de control.

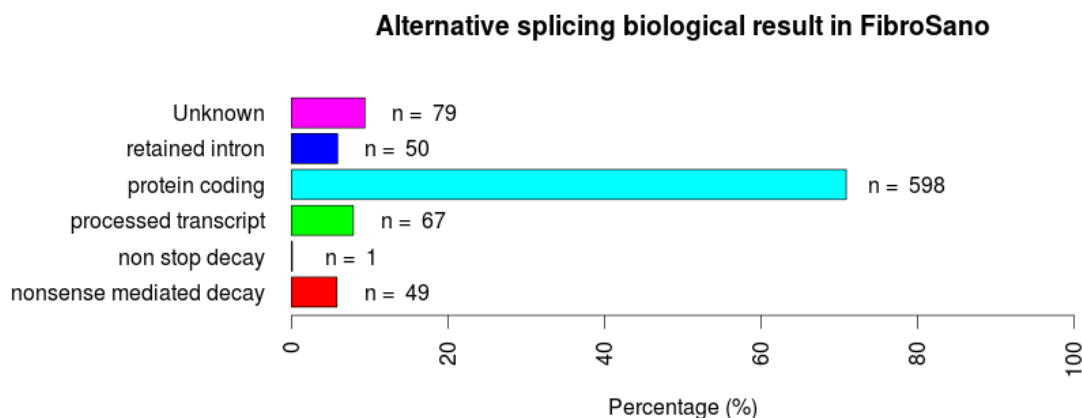
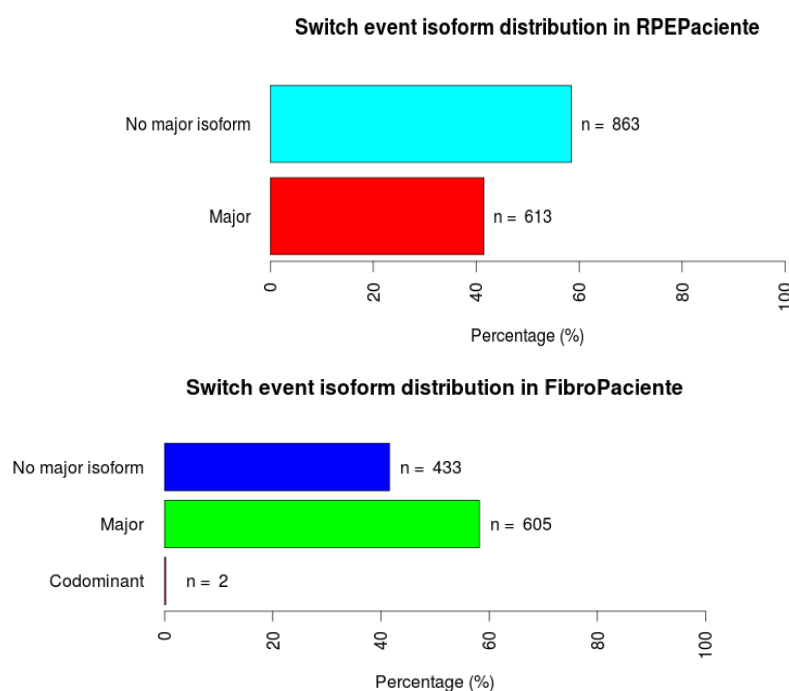


Fig. 18: Clasificación por biotipos de los transcritos que presentan switching que se asocian con la condición fibroblastos de control, en el contraste RPE de control Vs fibroblastos de control.

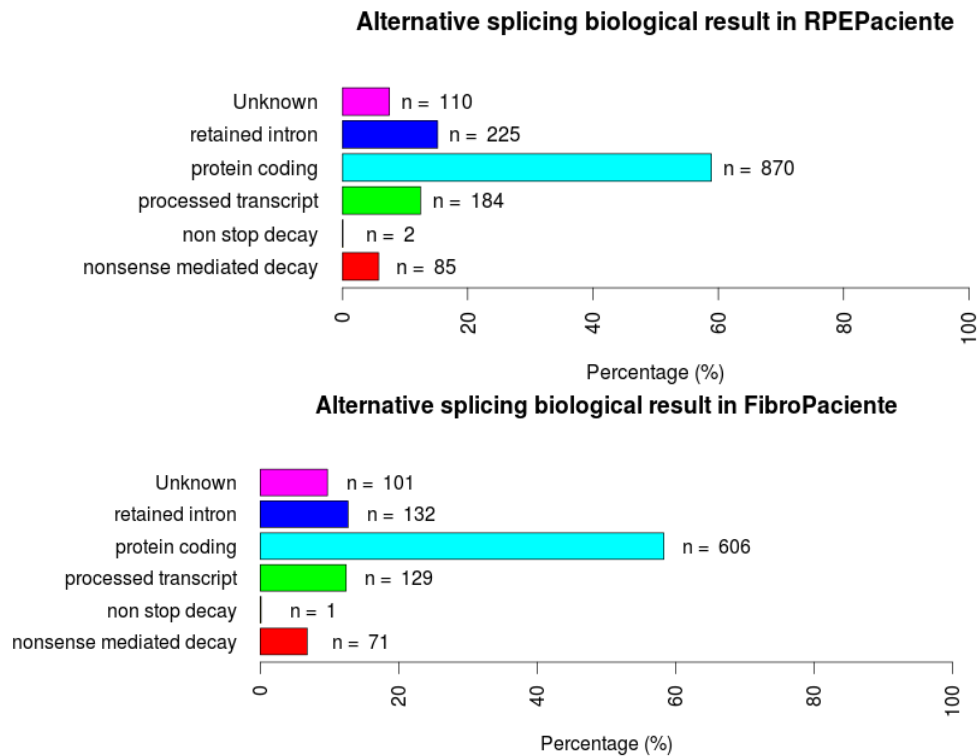
**En muestras de control, a diferencia de los dos contrastes anteriores, gran parte de los genes que experimentan switching, también tienen expresión diferencial. En RPE hay mayor numero de variantes transcripcionales de las que una alta proporción son no codificantes, especialmente con retención de intrón.**

#### 5.3.1.4. Comparación fibroblastos Vs RPE en paciente

En el paciente, al comparar la distribución de isoformas entre fibroblastos y células RPE, se detectaron 2516 variantes transcripcionales que cambiaban según condición, de las cuales 817 (33%) no estaban sujetas a expresión diferencial, figuras 19 y 20.



*Fig. 19: Distribución de variantes transcripcionales según abundancia relativa en células RPE de paciente (arriba) y en fibroblastos (abajo), en el contraste fibroblastos paciente Vs RPE de paciente.*



*Fig. 20: Clasificación por biotipos de los transcritos que presentan switching que se asocian con la condición células RPE de paciente (arriba) y la condición fibroblastos de paciente (abajo), en el contraste RPE de paciente Vs fibroblastos de paciente.*

***En pacientes, el perfil de splicing muestra menor número de variantes transcripcionales en células RPE Vs fibroblastos, que en control para el mismo contraste, lo cual demuestra la pérdida neta de transcritos con retención de intrón en RPE de paciente.***

### 5.3.2. RESULTADOS DE SwitchSeq

Los resultados para SwitchSeq refieren a casos de genes que presentan *switching* de isoformas y solo considera las 2 isoformas que más cambian, cada una de ellas asociada a una condición.

#### 5.3.2.1. Comparación paciente Vs control en células RPE.

Con SwitchSeq se detectaron 1794 casos de *switching*. En la figura 21, están representadas las distribuciones por biotipos de las isoformas sujetas a *switching*, en muestras de RPE de paciente y de control, respectivamente.

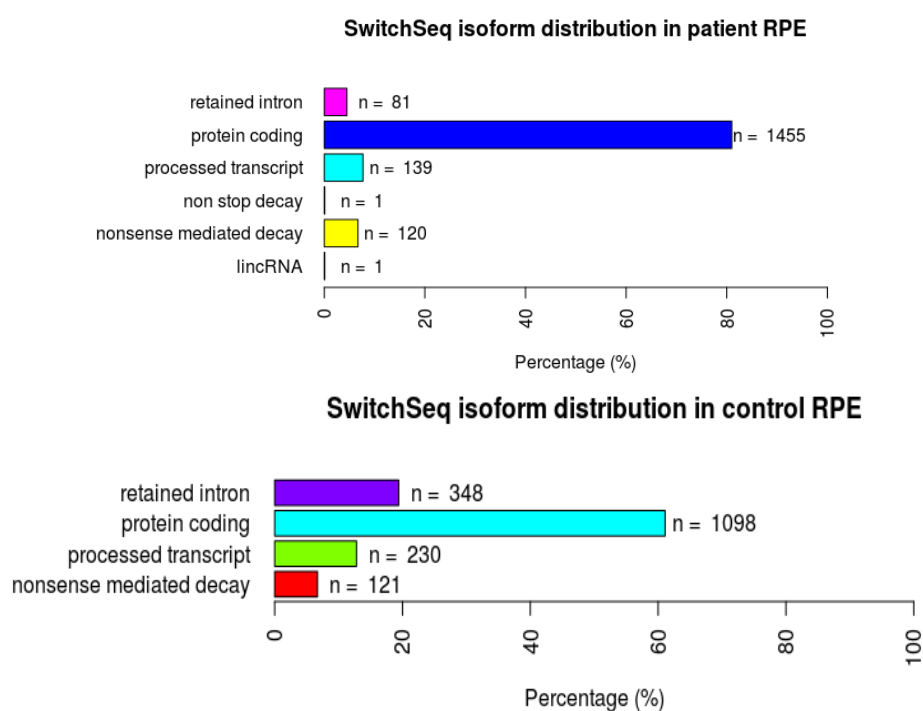
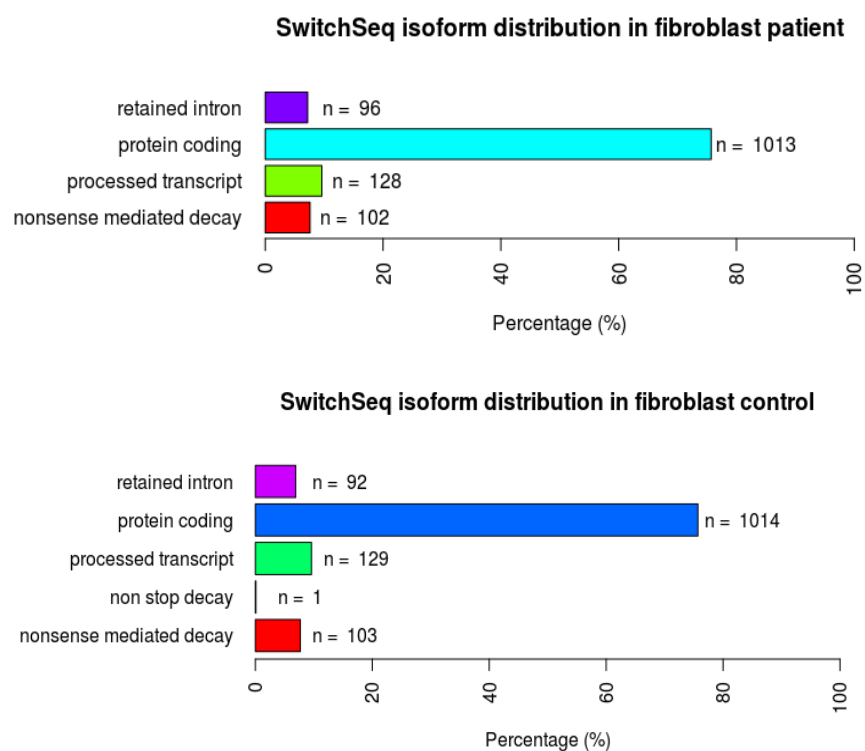


Fig. 21: Clasificación de SwitchSeq en biotipos de los transcritos asociados con la condición células RPE de paciente (arriba) y RPE de control(abajo), en el contraste RPE de paciente Vs RPE de control.

#### **5.3.2.2. Comparación paciente Vs control en fibroblastos.**

En fibroblastos, SwitchSeq reveló 1339 eventos de *switching* entre paciente y control. Sin embargo, como se puede observar en la figura 22, los perfiles de clasificación de las correspondientes isoformas en función del biotipo, resultaron prácticamente idénticos para fibroblastos de paciente y para fibroblastos de control.

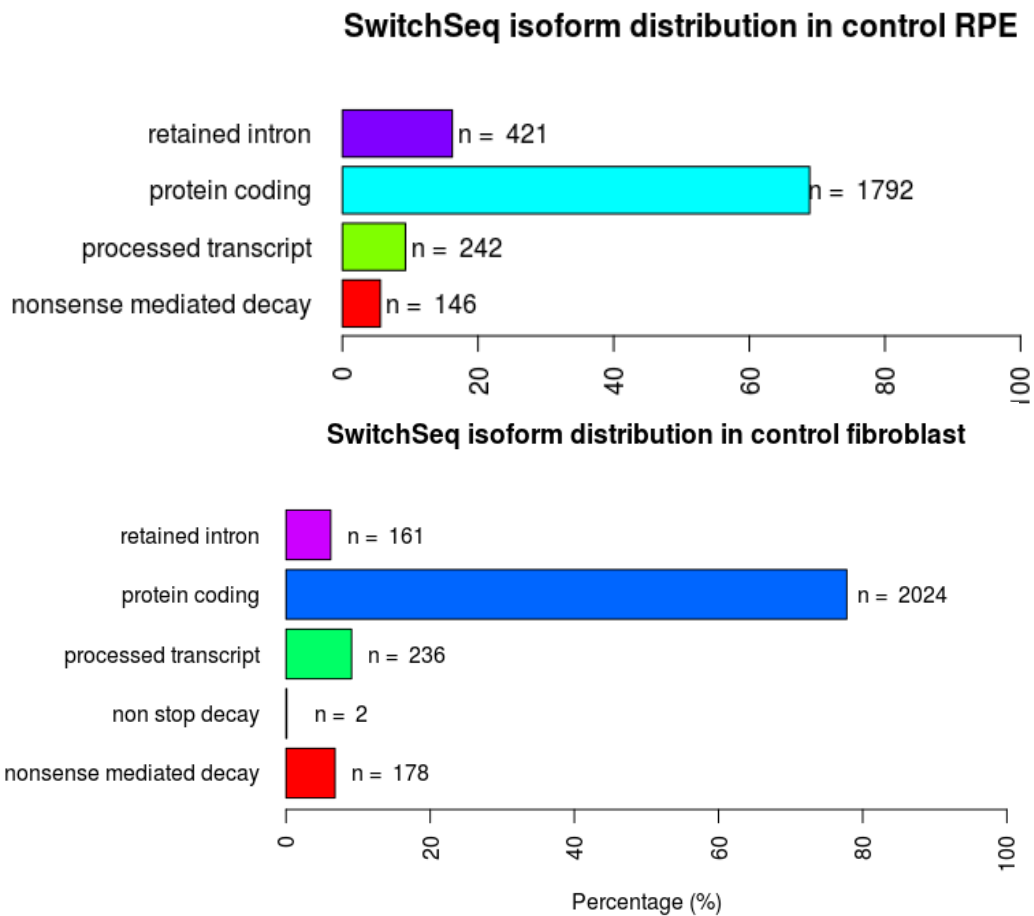




*Fig. 22: Clasificación de SwitchSeq en biotipos para los transcritos asociados con la condición fibroblastos de paciente (arriba) y la condición fibroblastos de control (abajo), en el contraste fibroblastos de paciente Vs fibroblastos de control.*

### 5.3.2.3. Comparación células RPE Vs fibroblastos en control.

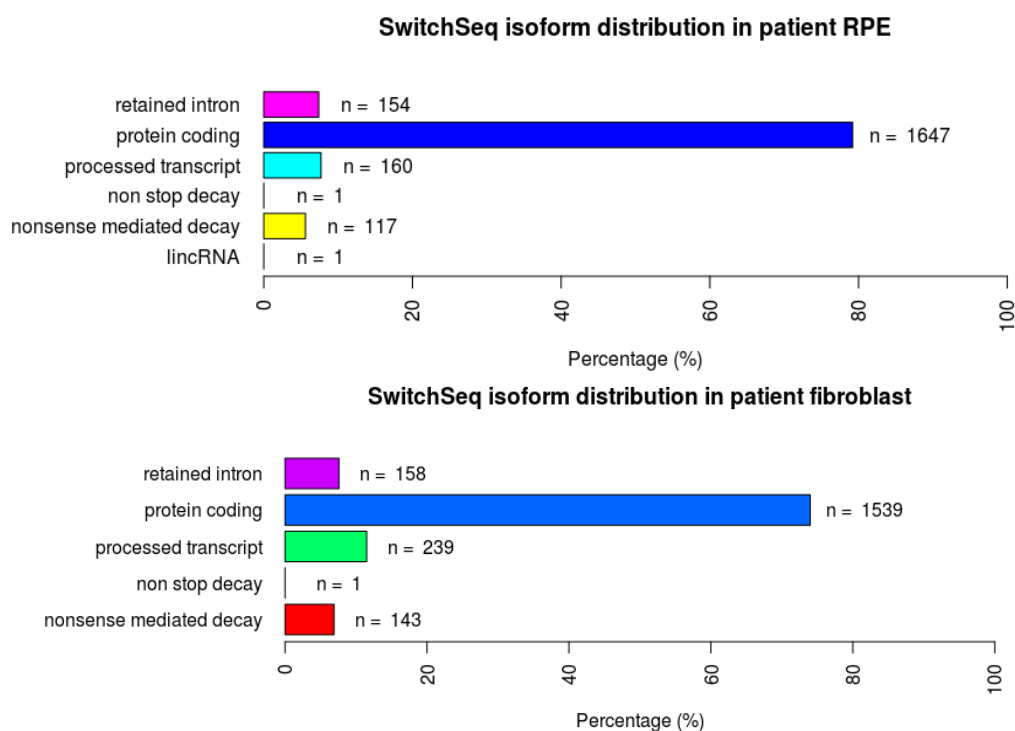
En muestras de control, 2601 genes mostraron *switching* de isoformas, figura 23.



*Fig. 23: Clasificación de SwitchSeq en biotipos para los transcritos asociados con la condición células RPE de control (arriba) y fibroblastos de control (abajo), en el contraste RPE de control Vs fibroblastos de control.*

#### **5.3.2.4. Comparación células RPE Vs fibroblastos en paciente.**

En muestras de paciente, se encontraron 2080 genes que presentaron *switching* de isoformas, figura 24.



*Fig. 24: Clasificación de SwitchSeq en biotipos para los transcritos asociados con la condición células RPE de paciente (arriba) y fibroblastos de paciente (abajo) en el contraste RPE de paciente Vs fibroblastos de paciente.*

***En resumen, lo presentado por SwitchSeq concuerda con los datos proporcionados por ASapp. En RPE de paciente se aprecia una disminución de las isoformas con retención de intrones, en relación a lo observado en muestras de control.***

### **5.3.3. RESULTADOS DEXSeq**

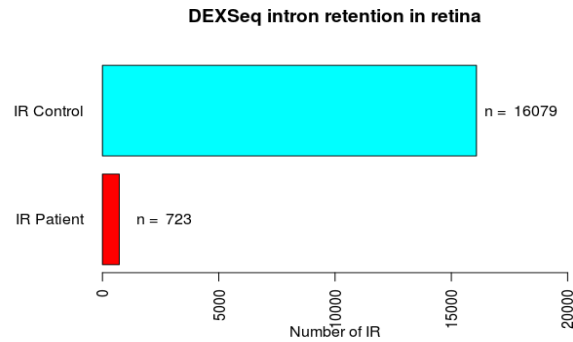
La herramienta DEXSeq está diseñada para la detección de eventos de uso diferencial de exones (DEU), lo cual dificulta la comparación a nivel de isoforma y la consiguiente interpretación biológica. El análisis se llevó a cabo

considerando un p-valor ajustado inferior a 0.05, por el método de Benjamini-Hotchberg. Modificando el protocolo de DEXSeq podemos valorar, no solo la utilización diferencial de exones, si no también la utilización diferencial de intrones. En el análisis incluimos agregados de genes, que consisten en genes que presentan solapamiento y se analizan conjuntamente, por consiguiente, aunque en lo sucesivo me refiera a genes, y en la mayoría de los casos sea cierto, también estaré incluyendo dentro de genes el concepto de estos agregados génicos.

#### **5.3.3.1. Comparación paciente Vs control en células RPE.**

En RPE, DEXSeq mostró 19722 eventos de utilización diferencial de exones y 16802 eventos de utilización diferencial de intrones. El número de genes (agregados génicos incluidos), que al menos presentan un exón o un intrón diferencialmente usado, fue de 10556. Mientras que el número de genes que presentaban al menos un evento de utilización diferencial de intrón fue de 8652 genes.

Ahondando más aún en los casos de retención de intrones, en la figura 25, están representados el número de intrones diferencialmente usados en cada condición. Podemos observar que la cantidad de intrones que se retienen en RPE de control (16079 eventos) es mucho mayor que en RPE de paciente (723 eventos). Estos números se mueven en rangos distintos a los de las otras herramientas porque no opera a nivel de isoforma. Esos 16079 eventos involucran a 9919 genes, mientras que los 723 eventos de utilización diferencial de intrones en RPE de pacientes en relación con muestras RPE de control, involucraron a 657 genes distintos. Dentro de esos 657 genes, 357 solaparon con la lista 8210 genes, por lo que para esos genes se da utilización diferencial de intrones en ambas direcciones del contraste.



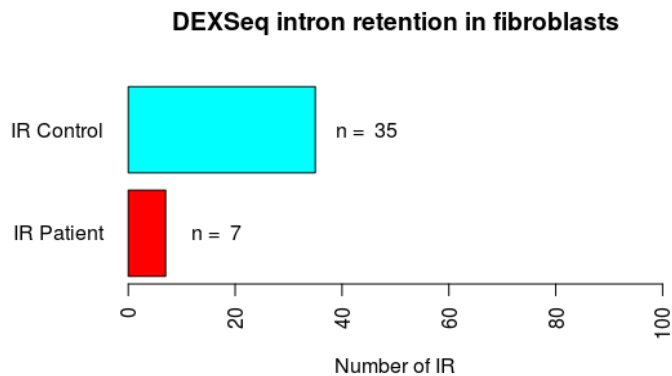
*Fig. 25: Resultados de DEXSeq para el uso diferencial de intrones (RPE de paciente Vs RPE de control).*

***Sintetizando los resultados de DEXSeq para células RPE, se confirma una alteración del patrón de splicing a consecuencia de la mutación. En la valoración del uso diferencial de intrones, para un  $FDR < 0.05$ , la cantidad de situaciones en las que un intrón se expresa de forma diferencial es muy superior en muestras del control, que en las del paciente.***

#### **5.3.3.2. Comparación paciente Vs control en fibroblastos.**

En fibroblastos se detectó uso diferencial de intrón/exón, entre paciente y control, en tan solo 154 unidades génicas, de las cuales en 41 se dio retención de intrón.

En fibroblastos, en concordancia con lo expuesto para datos de ASapp, la cantidad de retención de intrones fue mucho menor que en RPE. En control se dio retención de intrón en 42 ocasiones, frente a 7 casos de retención de intrón en paciente, figura 26. Cabe destacar que en este contraste el número de ocasiones en el que se da uso diferencial de intrones difiere ligeramente de lo expuesto por las otras herramientas, que resultan menos conservadoras.

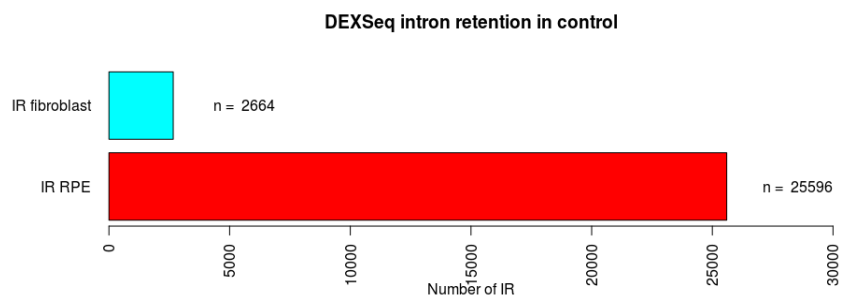


*Fig. 26: Resultados de DEXSeq para el uso diferencial de intrones (fibroblastos de paciente Vs fibroblastos de control).*

### 5.3.3.3. Comparación fibroblastos Vs células RPE en control.

Se observó utilización diferencial de elementos (intrón/exón) en 12208 unidades génicas, de los cuales en 10365 se dio utilización diferencial de intrones, para un valor de  $FDR < 0.05$ .

En la figura 27 se muestra la utilización diferencial de intrones entre fibroblastos y células RPE en el control. Del uso diferencial de intrones podemos determinar que en 25596 ocasiones un determinado intrón estuvo más utilizado en células RPE, con relación a fibroblastos, con 2664 retenciones de intrón.

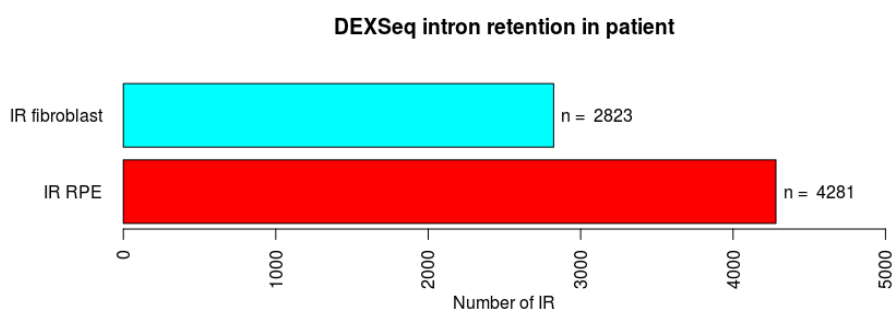


*Fig. 27: Resultados de DEXSeq para el uso diferencial de intrones (RPE de control Vs fibroblastos de control).*

**Los datos presentados permiten concluir, que la retención de intrón es marcadamente mayor en RPE que en fibroblastos.**

#### 5.3.3.4. Comparación fibroblastos Vs células RPE en paciente.

En pacientes se detectó utilización diferencial de intrón/exón en 8364 unidades génicas, de las cuales en 4465 se vio implicado al menos un intrón. En la figura 28 se muestra la utilización diferencial de intrones entre fibroblastos y células RPE. Podemos observar 4281 casos de retención de intrón en RPE, frente a 2823 retenciones de intrón en fibroblastos.



*Fig. 28: Resultados de DEXSeq para el uso diferencial de intrones (RPE de paciente Vs fibroblastos de paciente).*

***De esta forma, podemos concluir, que aunque la retención de intrones es mayor en RPE que en fibroblastos, es muy inferior a lo observado en el control.***

## **6. DISCUSIÓN**

En el presente estudio se analizan células RPE, que son células de elevada tasa metabólica, en constante actividad, de las que depende la supervivencia de las células fotorreceptoras. En consecuencia, disfunciones en células RPE se traducen en degeneración de bastones y de conos.

### **6.1 VALORACIÓN DE RESULTADOS DEL ANÁLISIS DE *SPLICING* ALTERNATIVO**

El *splicing* alternativo es un proceso altamente controlado y coordinado con la transcripción, que permite regular la expresión de los genes. Este fenómeno, está ampliamente extendido y está íntimamente asociado a la complejidad de los organismos (38).

Tradicionalmente, se considera al *splicing* como un mecanismo de diversificación, gracias al cual se alcanza una variabilidad mayor de transcritos susceptibles de ser traducidos. Estas variantes transcripcionales pueden diferir levemente o drásticamente, dando lugar a cambios en la eficacia del producto, variación en la localización o cambio de función (39)

En nuestro proyecto, los tres programas empleados para la valoración de AS, mostraron alteración del patrón de *splicing* en muestras del paciente. En el contraste entre células RPE de pacientes y RPE de control, las tres herramientas pusieron de manifiesto una reducción de isoformas con retención de intrones en muestras de paciente, lo cual no se observó en fibroblastos.

Cuando se valoraron las variaciones en los patrones de *splicing* en muestras de control, se vio que la proporción de isoformas con retención de intrones era muy superior en RPE que en fibroblastos. Esto también fue observado en muestras de paciente, sin embargo, el incremento de isoformas con retención de intrones en RPE fue mucho menor. Todo esto podría estar apuntando a un



mecanismo de regulación de la expresión, que operase a nivel de RPE, que a consecuencia de la mutación en PRPF8, se pierde. Esta modificación de los patrones normales de *splicing* implicaría una alteración en la expresión, que podría desencadenar la degeneración de las células RPE primero, de los bastones después y finalmente de los conos.

La proteína PRPF8 es un elemento indispensable dentro del complejo spliceosoma, que presenta un dominio RNasa H, responsable de la actividad catalítica en el *splicing*. Este dominio podría variar su actividad en torno a dos conformaciones, una de alta eficiencia y baja precisión y otra de baja eficiencia y alta precisión. De esta forma, diversas mutaciones podrían afectar a dichas conformaciones, bloqueando una u otra, generando patrones opuestos en el *splicing* (40). Esto explicaría en términos generales, el patrón de *splicing* observado en células RPE. De este modo, la delección de 21 pb, determina una pérdida del segmento C-terminal de la cadena, indispensable para la interacción con EFTUD2, componente de U5 snRNP. Por lo que, al alterarse la forma en que ambas proteínas interaccionan, se podría generar un cambio estructural tal, que quedase bloqueada la conformación de alta precisión/baja eficiencia del dominio RNasa H. En consecuencia, se aprecia una reducción general en la cantidad de isoformas distintas en RPE de paciente y una menor proporción de isoformas con retención de intrones.

El fenómeno de la retención de intrones se ha descrito, en muchos casos, como una fuente de variabilidad proteómica, una vez los transcritos escapan al mecanismo de conservación/control NMD (41). En nuestros datos, un ejemplo de ello es la isoforma ENST00000484980 de UNC119, donde la incorporación de un intrón, implica la aparición de un codón de parada que hace que la cadena polipeptídica codificada tenga 125 aa, en lugar de los 240 aa de la isoforma canónica.

Hasta la fecha, hay numerosos trabajos que demuestran, que la amplia diversidad observada a nivel de isoformas de transcritos, es mayor que la

diversidad demostrada en experimentos de proteómica (42). En este marco, los intrones, considerados durante mucho tiempo como “basura” remanente de ancestros arcaicos, podrían ocupar un papel indispensable en la regulación génica vía *splicing* (41,43). De hecho, su mera conservación evolutiva, por no mencionar el coste energético que el proceso de *splicing* en sí y su regulación representan, apuntan a la posibilidad de que desempeñen un papel relevante en este sentido. Por tanto, podría contribuir al complejo entramado de señalización que determina como, donde y en que proporción se expresan los productos génicos, elevando exponencialmente la complejidad de la regulación de la expresión fenotípica.

## **6.2. VALORACIÓN DE LOS RESULTADOS DEL ANÁLISIS DE EXPRESIÓN DIFERENCIAL Y FUNCIONAL.**

Como cabría esperar, dado que las células RPE de paciente presentan la mutación V2325fsX2329 para PRPF8, está representada de manera significativa la familia de genes implicada en el desarrollo de spliceosomas, el procesamiento de m-RNA y *splicing*. Más relevante es el hecho de que esté activada la ruta de la fosforilación oxidativa en paciente, dado que esto conlleva la formación de especies reactivas de oxígeno, que favorecen el estrés oxidativo.

Otro hallazgo relevante, es que estén representados genes que participan en la activación del sistema UPR y la exportación de proteínas. Cualquier disrupción en el funcionamiento normal del retículo endoplasmático, se traduce en estrés de este orgánulo y activación del sistema UPR. El sistema UPR, activa la expresión de chaperonas, reduce la traducción y promueve la eliminación de proteínas plegadas incorrectamente, en sentido de restituir la homeostasis. Por ello, también se encuentra activada la degradación de proteínas por la vía Ubiquitin/Proteasome System (UPS). De forma normal, en las células hay diversas vías de eliminación de proteínas, siendo el sistema UPS una de ellas. UPS participa en el turn-over normal de las

proteínas, dentro de procesos complejos como la fototransducción o la regulación del ciclo celular. Sin embargo, en retina, UPS también puede estar relacionado con fenómenos de estrés oxidativo, como mecanismo de defensa frente a la acumulación de proteínas oxidadas, que serían tóxicas y son responsables de la degeneración subsiguiente (44).

En consonancia con esto, en el análisis ORA de rutas KEGG, están representadas rutas responsables del desarrollo de Parkinson y de la enfermedad de Huntington. Estas patologías neurodegenerativas se relacionan con fenómenos de estrés oxidativo y la consiguiente activación del sistema de degradación de proteína UPS (44,45).

### **6.3 VALORACIÓN DE LA EXPRESIÓN DIFERENCIAL DE GENES DE LA LISTA RP**

El uso de técnicas de text mining permitió obtener una colección de genes relevantes para la RP, a partir de evidencias bibliográficas. Esto nos permitió evaluar el comportamiento individual de estos genes, a nivel de expresión, en el contexto de la enfermedad.

A excepción de RIMS1, ningún gen de la lista mostró expresión diferencial en fibroblastos, por tanto nos centraremos en el contraste entre células RPE de paciente y RPE de control.

A pesar de ser interesante valorar la presencia de *switching* dentro de estos genes, la integración de la información de los tres programas no es una tarea trivial, por lo que queda pendiente para futuros trabajos.

#### **6.3.1. Genes de la lista RP sobre-expresados**

La mayoría de los genes de la lista RP son importantes en caso de pérdida de función de los mismos, por lo que la sobre-expresión de estos genes, puede no tener relevancia, a priori, en la patogénesis. Además, puede que algunos de esos genes sean desencadenantes de la enfermedad en el

contexto de otras células de la retina, como los fotorreceptores o las neuronas bipolares, por lo que su comportamiento en RPE, salvo excepción, solo representa una aproximación. En consecuencia, en este apartado, discutiremos solo aquellos casos para los que la activación en RPE de paciente sea relevante, frente a muestras RPE de control. Dentro de este grupo de genes, los que estaban sobre-expresados de manera significativa en RPE de paciente fueron C2, CFB, KHL7, TOPORS, ATF6, EIF2AK3, OPTN y MAPT.

En el caso de C2 y CFB, ambos componentes del sistema de complemento, su activación responde a la sobre-expresión de IL6. A este estado pro-inflamatorio generado por IL6, contribuye la activación adicional del grupo de quimioquinas CXC y la proteína C1qL1. Todo ello conduciría a una eventual activación del sistema de complemento que favorecería la degeneración de las células. Hallazgos que concuerdan con trabajos que señalan al estado crónico de inflamación como responsable de la patogénesis de RP (46). Análogamente, se detectó activación para CD59 en RPE de paciente, que es un inhibidor del complejo de ataque del sistema de complemento. De esta forma el estrés oxidativo conduce al desarrollo de un estado inflamatorio crónico, que activa el sistema de complemento y del que las células intentan protegerse aumentando la expresión de CD59 (47).

KHL7 y TOPORS son genes que participan en la degradación de proteínas mediante el sistema UPS, descrito más arriba. Por lo que también se activarían en respuesta a estrés.

Por su parte, ATF6 y EIF2AK3 (ambos integrantes de UPR) son dos agentes que participan en la inhibición de síntesis de proteínas, como respuesta a situaciones de estrés en el retículo endoplasmático.

Finalmente, la acumulación en el citoplasma de la optoneurina y la proteína tau, codificadas respectivamente por OPTN y MAPT, podría estar relacionada

con la fragmentación del aparato de Golgi, hecho que podría contribuir a la degeneración celular (48,49).

### **6.3.2. Genes de la lista RP reprimidos**

En el capítulo de genes reprimidos en RPE de paciente respecto a RPE de control, hay gran cantidad de genes de relevancia estructural y funcional para el desarrollo de RP como CA4, FLVCR1, NPHP1, NPHP4, CEP290, SPATA7, EFHC1, WHRN, IFT172, WDR73, WDR19, OFD1, KHDRBS2 y PRPF3.

La proteína codificada por CA4 es una anhidrasa carbónica, que se encarga de procesos homeostáticos en retina, concretamente del mantenimiento del pH. Se han descrito mutaciones en este gen responsables de la degeneración de la retina (50). Por tanto, su represión, aunque no fuera absoluta, se traduciría en una disminución del pH que contribuiría a la degeneración de la célula RPE.

En este mismo sentido, el gen FLVCR1, codifica para un transportador del grupo hemo, cuyo metabolismo está críticamente involucrado en el mantenimiento de neuronas sensoriales y células RPE. En consecuencia una reducción de la función de FLVCR1, implicaría la acumulación tóxica del grupo hemo(51).

Los cilios son unas estructuras celulares indispensables en células fotorreceptoras. En términos generales, la disfunción ciliar es la causa de gran número de enfermedades, que cursan con degeneración de la retina. Por lo que alteraciones en la ciliogénesis, el mantenimiento de la estructura ciliar y el transporte a través de cilios, conllevan el desarrollo de RP. Se ha descrito que las células RPE en cultivo desarrollan cilios primarios, en consecuencia son modelos adecuados para el estudio de la ciliogénesis(52).

Dentro de los genes reprimidos en RPE de pacientes, hay una serie de genes que contribuyen a la formación y mantenimiento de la estructura ciliar del

fotorreceptor y a la determinación de la polarización celular, como el grupo de las Nefrocistinas NPHP1, NPHP4 y CEP290 (también llamada NPHP6), SPATA7, EFHC1 y WHRN. IFT172, WDR73 y WDR19 (también llamado FT144) son genes implicados en el transporte intraciliar en el fotorreceptor. Todos ellos son genes para los que se han descrito mutaciones que conllevan el desarrollo de RP (53–57).

OFD1, codifica para un componente de los centriolos, que regula la longitud y el número de estructuras ciliares, es también un neuro-protector que preserva a las células de la retina frente a su degeneración, protegiéndolas del estrés oxidativo y de la apoptosis (58). Este gen también está relacionado con el desarrollo de RP (59). De modo que la represión de este gen, puede estar implicada de manera directa en la degeneración de la célula, al encontrarse reprimida diferencialmente.

KHDRBS2 y PRPF3 son constituyentes de la maquinaria celular destinada a procesos de *splicing*, para los que también se han descrito mutaciones responsables del desarrollo de RP (5). Su represión parcial podría contribuir a la pérdida de control sobre los mecanismos de *splicing*.

En resumen, la alteración de los patrones normales de *splicing*, conlleva una alteración de la expresión proteica en células RPE. Ésto determina el mal plegamiento de proteínas a nivel del retículo endoplásmico y la posterior activación del sistema UPR (donde se integra ATF6). La activación UPR, a su vez conduce a la activación del sistema UPS de degradación de proteínas, en el sentido de restablecer la homeostasis. Paralelamente, el estrés a nivel del retículo endoplásmico determina la aparición de un estado proinflamatorio, con activación del sistema de complemento, frente al que las células RPE se protegen mediante la sobre-expresión de CD59. Al encontrarse detenida la traducción, se ve afectada la exportación, a lo que contribuye la acumulación de OPTN y MAPT, que favorece la fragmentación del aparato de Golgi y el funcionamiento normal del mismo.

A pesar de que la ciliogénesis es un proceso que tiene relevancia en fotorreceptores, las células RPE, al presentar ciliogénesis en cultivo, constituyen una aproximación razonable de lo que pasaría en conos y bastones. Por eso, la represión observada para proteínas importantes en la génesis, mantenimiento y transporte a través de estas estructuras en RPE, podría ser la situación observada en fotorreceptores.

Adicionalmente, las células RPE, en su entorno fisiológico normal, están sometidas a fenómenos de foto-oxidación, principalmente debido a la luz azul, dado que la UV es absorbida de forma eficaz por filtros previos. La foto-oxidación determina la aparición de más especies reactivas del oxígeno que agravan la situación descrita.

Todas estas alteraciones se integrarían en una enfermedad compleja, de carácter degenerativo, cuyas bases genéticas son variadas, que hemos de concebir como el resultado acumulativo del incorrecto funcionamiento de diversos procesos celulares. Sin embargo todo ello puede ser disparado por una única mutación, en este caso por la mutación V2325fsX2329 para PRPF8.

#### **6.4. VALORACIÓN COMPARADA DE ASAPP, DEXSEQ Y SWITCHSEQ**

Los tres programas mostraron resultados similares pese a que sus estrategias de análisis son diferentes. Los resultados en SwitchSeq son más moderados en sus números, probablemente porque solo proporciona un transcrito por condición. A falta de un análisis más exhaustivo con datos simulados o algún *benchmark* válido para el análisis de splicing alternativo, que nos proporcione una medida del error cometido por ASapp, podemos concluir que ASapp es un programa útil para la valoración e interpretación biológica del splicing alternativo.

## **7. CONCLUSIONES:**

1. La mutación V2325fsX2329 en el gen PRPF8, produce el desajuste del *splicing* normal en RPE de paciente, con una marcada disminución de isoformas con retención de intrones.
2. Debido a la modificación de los perfiles normales de *splicing*, se ve alterado el perfil de expresión en RPE de paciente.
3. La alteración en el perfil de expresión conduce a una situación de estrés oxidativo, con pérdida del correcto funcionamiento del RE, activación de los sistemas UPR y UPS, inflamación y activación del sistema de complemento.
4. La ciliogénesis, mantenimiento y transporte a través de las estructuras ciliares se desajusta, por cambios en la expresión normal de agentes implicados.
5. Los eventos anteriormente mencionados conducen a la disfunción de las células RPE, lo cual puede desencadenar la degeneración primaria de los bastones.
6. ASapp es una herramienta útil para el análisis de *splicing* alternativo.



## **8. BIBLIOGRAFÍA:**

1. Ali MU, Rahman MSU, Cao J, Yuan PX. Genetic characterization and disease mechanism of retinitis pigmentosa; current scenario. 3 Biotech. 2017;7(4).
2. Steinberg RH. Interactions between the retinal pigment epithelium and the neural retina. Doc Ophthalmol [Internet]. 1985;60(4):327–46. Available from: <http://link.springer.com/10.1007/BF00158922>
3. Streilein JW, Ma N, Wenkel H, Fong Ng T, Zamiri P. Immunobiology and privilege of neuronal retina and pigment epithelium transplants. Vision Res. 2002;42(4):487–95.
4. Martínez-Gimeno M, José Gamundi M, Hernan I, Maseras M, Millá E, Ayuso C, et al. Mutations in the pre-mRNA splicing-factor genes PRPF3, PRPF8, and PRPF31 in Spanish families with autosomal dominant retinitis pigmentosa. Invest Ophthalmol Vis Sci. 2003;44(5):2171–7.
5. Gonzalez-Santos JM, Cao H, Duan RC, Hu J. Mutation in the splicing factor Hprp3p linked to retinitis pigmentosa impairs interactions within the U4/U6 snRNP complex. Hum Mol Genet. 2008;17(2):225–39.
6. Neveling K, Collin RWJ, Gilissen C, Van Huet RAC, Visser L, Kwint MP, et al. Next-generation genetic testing for retinitis pigmentosa. Hum Mutat. 2012;33(6):963–72.
7. Matlin AJ, Clark F, Smith CWJ. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol [Internet]. 2005;6(5):386–98. Available from: <http://www.nature.com/doifinder/10.1038/nrm1645>
8. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. Gene [Internet]. 2013 Feb;514(1):1–30. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0378111912009791>
9. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature

- [Internet]. 2008;456(7221):470–6. Available from:  
<http://www.nature.com/doifinder/10.1038/nature07509>
10. Moore MJ, Wang Q, Kennedy CJ, Silver PA. An Alternative Splicing Network Links Cell-Cycle Control to Apoptosis. *Cell* [Internet]. 2010 Aug;142(4):625–36. Available from:  
<http://linkinghub.elsevier.com/retrieve/pii/S0092867410007889>
  11. Li Y, Tsai Y-T. Long-term safety and efficacy of human-induced pluripotent stem cell (iPS) grafts in a preclinical model of retinitis pigmentosa. *Mol Med* [Internet]. 2012;18(9):1. Available from:  
[http://www.molmed.org/pdfstore/12\\_242\\_Li.pdf](http://www.molmed.org/pdfstore/12_242_Li.pdf)
  12. Lukovic D, Bolinches Amoros A, Castro AA, Pascual B, Carballo M, Hern??n I, et al. Generation of a human iPSC line from a patient with retinitis pigmentosa caused by mutation in PRPF8 gene. *Stem Cell Res* [Internet]. 2017;21:23–5. Available from:  
<http://dx.doi.org/10.1016/j.scr.2017.03.007>
  13. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* [Internet]. 2009 Jan;10(1):57–63. Available from: <http://www.nature.com/doifinder/10.1038/nrg2484>
  14. Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* [Internet]. 2012;22(4):271–4. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3426205&tool=pmcentrez&rendertype=abstract>
  15. Pachter L. Models for transcript quantification from RNA-Seq. 2011 Apr 19; Available from: <http://arxiv.org/abs/1104.3889>
  16. Rodriguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink JJ, Lopez G, et al. APPRIS: Annotation of principal and alternative splice isoforms. *Nucleic Acids Res*. 2013;41(D1):110–7.

17. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. Available from:  
<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
18. Martin. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. [Internet]. EMBnet.journal, 17(1); 2011. p. 10–2. Available from:  
<http://journal.embnnet.org/index.php/embnnetjournal/article/view/200>
19. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics [Internet]. 2011;12(1):323. Available from:  
<http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-323>
20. Li, Bo & Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. [Internet]. BioMed Central: The Open Acces Publisher. Retrieved at October 10; 2011. Available from:  
<http://www.temoa.info/node/441614>
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
22. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol [Internet]. 2014;15(2):R29. Available from:  
<http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r29>
23. Phipson B, Lee S, Majewski IJ, Alexander WS SG. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression 1. Ann Appl Stat [Internet]. 2016;10(2):946–63. Available from:  
[https://projecteuclid.org/download/pdfview\\_1/euclid.aoas/1469199900](https://projecteuclid.org/download/pdfview_1/euclid.aoas/1469199900)

24. Carbon S, Dietze H, Lewis SE, Mungall CJ, Munoz-Torres MC, Basu S, et al. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Res.* 2017;45(D1):D331–8.
25. Schriml LM, Arze C, Nadendla S, Chang YWW, Mazaitis M, Felix V, et al. Disease ontology: A backbone for disease semantic integration. *Nucleic Acids Res.* 2012;40(D1):940–6.
26. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457–62.
27. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi A J Integr Biol* [Internet]. 2012 May;16(5):284–7. Available from: <http://online.liebertpub.com/doi/abs/10.1089/omi.2011.0118>
28. Subramanian, Tamayo et al. Gene set enrichment analysis, GSEA software, and Molecular Signature Database (MSigDB) [Internet]. *PNAS* 102; 2005. p. 15545–50. Available from: <http://www.broad.mit.edu/gsea/>
29. Fontaine JF, Priller F B-SA and A-NM. Génie: literature-based gene prioritization at multi genomic scale. *Nucleic Acids Research* 2011 Jul;39(Web Server issue):W455-61. 2011.
30. Gonzàlez-Porta M, Brazma A. Identification, annotation and visualisation of extreme changes in splicing from RNA-seq experiments with SwitchSeq. Pre-Print [Internet]. 2014;10.1101/005967. Available from: <http://biorxiv.org/content/early/2014/06/06/005967.abstract>
31. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 2012;22(10):2008–17.
32. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* [Internet]. 2013;14(4):R36. Available from: <http://genomebiology.com/2013/14/4/R36>

33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
34. Winston Chang, Joe Cheng, JJ Allaire YX and JM. shiny: Web Application Framework for R. [Internet]. 2017. Available from: <https://cran.r-project.org/package=shiny>
35. Hu Y, Huang Y, Du Y, Orellana CF, Singh D, Johnson AR, et al. DiffSplice: The genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res*. 2013;41(2):1–18.
36. Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat Biotechnol* [Internet]. 2011;28(5):511–5. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3146043&tool=pmcentrez&rendertype=abstract>
37. Singh D, Orellana CF, Hu Y, Jones CD, Liu Y, Chiang DY, et al. FDM: A graph-based statistical method to detect differential transcription using RNA-seq data. *Bioinformatics*. 2011;27(19):2633–40.
38. Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. Correcting for Differential Transcript Coverage Reveals a Strong Relationship between Alternative Splicing and Organism Complexity. *Mol Biol Evol* [Internet]. 2014 Jun;31(6):1402–13. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu083>
39. Wright KJ, Baye LM, Olivier-Mason A, Mukhopadhyay S, Sang L, Kwong M, et al. An ARL3-UNC119-RP2 GTPase cycle targets myristoylated NPHP3 to the primary cilium. *Genes Dev*. 2011;25(22):2347–60.
40. Mayerle M, Guthrie C. Prp8 retinitis pigmentosa mutants cause defects in the transition between the catalytic steps of splicing. *RNA* [Internet]. 2016;rna.055459.115-. Available from:

- <http://rnajournal.cshlp.org/content/early/2016/03/11/rna.055459.115.abstract?sid=2ec447b3-a88f-4c86-825e-dda9011fd06c>  
%5Cn<http://www.ncbi.nlm.nih.gov/pubmed/26968627>
41. Hubé F, Francastel C. Mammalian Introns: When the Junk Generates Molecular Diversity. *Int J Mol Sci* [Internet]. 2015 Feb 20;16(3):4429–52. Available from: <http://www.mdpi.com/1422-0067/16/3/4429/>
  42. Ezkurdia I, Del Pozo A, Frankish A, Rodriguez JM, Harrow J, Ashman K, et al. Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Mol Biol Evol*. 2012;29(9):2265–83.
  43. McGlincy NJ, Smith CWJ. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem Sci*. 2008;33(8):385–93.
  44. Campello L, Esteve-Rudd J, Cuenca N, Martín-Nieto J. The Ubiquitin-Proteasome System in Retinal Health and Disease. *Mol Neurobiol*. 2013;1–21.
  45. Li J, O W, Li W, Jiang Z-G, Ghanbari H. Oxidative Stress and Neurodegenerative Disorders. *Int J Mol Sci* [Internet]. 2013;14(12):24438–75. Available from: <http://www.mdpi.com/1422-0067/14/12/24438/>
  46. Yoshida N, Ikeda Y, Notomi S, Ishikawa K, Murakami Y, Hisatomi T, et al. Clinical Evidence of Sustained Chronic Inflammatory Reaction in Retinitis Pigmentosa. *Ophthalmology* [Internet]. 2013 Jan;120(1):100–5. Available from: <http://dx.doi.org/10.1016/j.ophtha.2012.07.006>
  47. Yang P, Tyrrell J, Han I, Jaffe GJ. Expression and modulation of RPE cell membrane complement regulatory proteins. *Invest Ophthalmol Vis Sci*. 2009;50(7):3473–81.
  48. Park B-C, Shen X, Samaraweera M, Yue BYJT. Studies of Optineurin, a Glaucoma Gene. *Am J Pathol* [Internet]. 2006 Dec;169(6):1976–89. Available from:

- <http://linkinghub.elsevier.com/retrieve/pii/S0002944010626594%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0002944010626594%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1762487/pdf/7054.pdf>
49. Liazoghli D, Perreault S, Micheva KD, Desjardins M, Leclerc N. Fragmentation of the Golgi Apparatus Induced by the Overexpression of Wild-Type and Mutant Human Tau Forms in Neurons. *Am J Pathol* [Internet]. 2005;166(5):1499–514. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0002944010623668>
  50. Yang Z, Alvarez B V, Chakarova C, Jiang L, Karan G, Li X, et al. Mutant carbonic anhydrase 4 impairs pH regulation and causes retinal photoreceptor degeneration. 2017;14(2):255–65.
  51. Chiabrando D, Castori M, di Rocco M, Ungelenk M, Gießelmann S, Di Capua M, et al. Mutations in the Heme Exporter FLVCR1 Cause Sensory Neurodegeneration with Loss of Pain Perception. *PLoS Genet*. 2016;12(12):1–18.
  52. Kim J, Lee JE, Heynen-Genel S, Suyama E, Ono K, Lee K, et al. Functional genomic screen for modulators of ciliogenesis and cilium length. *Nature* [Internet]. 2010 Apr 15;464(7291):1048–51. Available from: <http://www.nature.com/doifinder/10.1038/nature08895>
  53. Ebermann I, Scholl HPN, Charbel Issa P, Becirovic E, Lamprecht J, Jurklics B, et al. A novel gene for Usher syndrome type 2: Mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Hum Genet*. 2007;121(2):203–11.
  54. Sayer JA, Otto EA, O'Toole JF, Nurnberg G, Kennedy MA, Becker C, et al. The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. *Nat Genet* [Internet]. 2006;38(6):674–81. Available from: <http://www.nature.com/doifinder/10.1038/ng1786>

- nephronophthisis-associated proteins in mammalian retinas. *Mol Vis* [Internet]. 2010;16(July):1373–81. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2905641&tool=pmcentrez&rendertype=abstract>
56. Otto E, Hoefele J, Ruf R, Mueller AM, Hiller KS, Wolf MTF, et al. A gene mutated in nephronophthisis and retinitis pigmentosa encodes a novel protein, nephroretinin, conserved in evolution. *Am J Hum Genet* [Internet]. 2002;71(5):1161–7. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=385091&tool=pmcentrez&rendertype=abstract>
57. Zhao Y, Shi J, Winey M, Klymkowsky MW. Identifying domains of EFHC1 involved in ciliary localization, ciliogenesis, and the regulation of Wnt signaling. *Dev Biol* [Internet]. 2016;411(2):257–65. Available from: <http://dx.doi.org/10.1016/j.ydbio.2016.01.004>
58. Wang J, Chen X, Wang F, Zhang J, Li P, Li Z, et al. OFD1, as a ciliary protein, exhibits neuroprotective function in photoreceptor degeneration models. *PLoS One*. 2016;11(5):1–18.
59. Webb TR, Parfitt DA, Gardner JC, Martinez A, Bevilacqua D, Davidson AE, et al. Deep intronic mutation in *ofd1*, identified by targeted genomic next-generation sequencing, causes a severe form of x-linked retinitis pigmentosa (rp23). *Hum Mol Genet*. 2012;21(16):3647–54.